心理科学进展 2022, Vol. 30, No. 4, 738–760
Advances in Psychological Science

https://doi.org/10.3724/SP.J.1042.2022.00738

·元分析(Meta-Analysis)·

自闭症谱系障碍的早期筛查工具*

陈光华[1] 陶冠澎[1] 翟璐煜[1] 白学军[2]

([1]沈阳师范大学学前与初等教育学院, 沈阳 110034) ([2]天津师范大学心理与行为研究院, 天津 300074)

摘 要 基于35项实证研究, 遵循诊断准确性研究的质量评估(QUADAS-2)的循证程序对自闭症谱系早期预警特征筛查工具开展质量评估, 包括婴儿阶段(10个筛查工具, 159388人)、幼儿阶段(14项筛查工具, 11712人), 旨在为自闭症谱系风险的早期识别提供全面的理解。结果表明: 幼儿阶段的筛查工具的分类准确性要高于婴儿阶段, 在婴儿阶段达到良好水平的筛查工具有M-CHAT-R/F、PDQ-1; 在幼儿阶段达到优秀水平的筛查工具有OERA、TIDOS。其中, "改良的幼儿自闭症检查表−有修订的后续随访" (M-CHAT-R/F)是目前最具潜力的自闭症谱系风险筛查工具之一。最后, 我们探讨了应用QUADAS-2评估研究质量的局限性, 强调需要规范工具的质量评估标准与进一步验证研究的必要性。

关键词 自闭症, 筛查工具, 诊断敏感性和特异性, QUADAS-2

分类号 R395

1 引言

筛查(screening)或筛查评估是"一种相对简短的评估, 旨在确定有可能患上某些疾病或残疾的儿童、有资格参加某些项目的儿童、有需要补救的疾病或残疾的儿童或需要更全面评估的儿童" (Sattler, 2008)。检视以往的中文文献, 我们发现, 关于自闭症谱系早期筛查的研究有两方面的误区, 第一个误区表现在对于"筛查"的概念和操作规范的认知不足, 常将之与"诊断"和"评估"放在一起(张福娟, 贺莉, 2001), 没有明确区分"筛查"作为一个独立而重要的阶段来开展广泛深入性的研究。首先, 我们根据《特殊教育辞典》的概念界定(朴永馨, 2014)来说明三者的区别。"筛查(screening)"是在大量儿童中发现特殊儿童(主要是残疾儿童)的一种检测活动。"诊断(diagnosis)"是对筛查出的特殊儿童或怀疑为特殊的儿童进一步做全面细致的专业检查。"鉴定(assessment)"又译为"评估"、"评定", 是由相应的机构或专业人员用一定的方法对疑似特殊儿童进行各种必要检查并做出结论的过程。其次, 从操作流程上, 筛查发生在正规诊断之前, 筛查前的转介可能来自产前检查 、新生儿筛选、儿童保健门诊、社区护士、医院门诊、公立或民办幼儿园、学前教育机构或托儿所、儿童福利机构及托育中心、家长或监护人、保姆等多元化的通报(何华国, 2006), 筛查是为了判断某儿童是或不是某类特殊需要儿童(王辉, 2015), 虽然自闭症筛查的结论只能是该儿童不是自闭症或者可能是自闭症, 但可以有效地提高之后自闭症诊断的准确性和可靠性。正规诊断是由受过训练取得合格证书的专业人员进行, 使用的测查工具更复杂, 操作程序更严格和标准化。诊断结果应有书面结论, 包括特殊儿童的类型、性质、程度、诊断时间、地点、测查人员签名和测查单位盖章。而评估是一个多层面的过程, 这个术语在特殊教育领域中应用更为宽泛, 特殊儿童的评估一般由特殊教育学、特殊心理学、有关医学等方面专家和教育行政部门代表、教师以及家长共同参加。评估过程通常包含如下目的: 筛选、诊断与安置、帮助制定个别教育计划, 帮

收稿日期: 2021-07-15
* 2020年辽宁省社会科学基金一般项目: 自闭症儿童的家庭照顾和社会支持体系研究(L20BSH010)。
通信作者: 陈光华, E-mail: ghse@163.com;
白学军, E-mail: bxuejun@126.com

助确定学生的进步情况并判断有关教育干预措施的有效性(方俊明, 2005)。

另一个误区是，对筛查工具开发和使用目的的理解不到位。例如，有一部分研究者混淆了“筛查工具”和“诊断工具”的界限，将金标准意义的ADI-R 和 ADOS 这两项“临床诊断工具”简单地等同于“筛查工具” (张永盛，吕超，2013; 冯雅静，王雁, 2012)。筛查工具使用的重点在于其“简短”、“经济”、“快捷”、“准确”地早期识别目的和功能，而不能被看作是 ASD 的诊断工具，因为正规的诊断必须严格遵循 DSM 或 ICD 的诊断标准，只有执业医师在必要的临床环境中，借助“ADI-R”和“ADOS”的临床诊断工具的同时，还要进行大量的身体和行为测试，才有资格对自闭症做出正规诊断。经过正规诊断的孩子才有资格获得政府和保险机构的支持和资源，进而安排早期干预治疗和教育安置工作。在这个角度来看，ASD 的诊断是非常复杂、主观、耗时且昂贵的，但最为关键的是，筛查过程需要由后续的诊断结果来评判筛查工具的准确性，正如许多研究者所建议的(Falkmer et al., 2013; Volkmar et al., 2014), “早期筛查的过程应该包括早期筛查预警信号和通过临床判断做出的后续诊断，并结合应用可靠和标准化的金标准测量，例如，ADI-R; ADOS-2”。因此，诊断工具与筛查工具不能相互替代。另外，还有部分研究者将“筛查工具”等同于“评估工具” (王辉 等, 2009)、“测评工具” (周念丽，方俊明, 2008), 首先，这一误解在于忽视了筛查工具在临床诊断和教育评估中的重要意义。它不仅可以帮助家长和从业者尽早识别自闭症风险行为，加快临床诊断转诊流程，而且有助于提高不同利益相关者(父母、照顾者、教师和家庭成员等)对自闭症的理解，改善公众对自闭症的认知。其次，这一误解在于没有辨识到早发现与早诊断之间的渐近性，这对于避免诊断延迟，尽早干预至关重要。美国父母通常会在 24 个月前发现并表示孩子可能患有自闭症的担忧(CDC, 2012), 中国家长注意到症状并寻求诊断的时间平均在 35 个月(Zhou et al., 2014)，然而，只有 4 岁甚至 8 岁以后才能确诊(CDC, 2014), 在有色人种儿童和农村及低收入家庭儿童的诊断年龄甚至更晚(Mandell et al., 2010)。为了摆脱诊断延迟的困境，保健工作者和其他专业相关人员重点关注了定期使用发育筛查的效果，而且有增加的证据表明，筛查的确会降低诊断年龄。如 Carbone 等人(2020)发现，早期筛查阳性的儿童被诊断 ASD 的时间要比筛查阴性的儿童早 12 个月，比未参加筛查的儿童早 10 个月。因此，开展自闭症谱系早期筛查目的在于尽早发现疑似自闭症谱系症状的婴幼儿，通过接受基本的测试，以确定其是否需要进一步的诊断评估和早期干预。通过这一系列渐进性的早识别，早诊断，早干预过程不仅可以增强自闭症谱系障碍儿童的适应能力(Dawson et al., 2010), 提高他们的社交能力(Leo et al., 2019), 预防继发性发育障碍(Reichow et al, 2012), 更有助于减轻家庭的压力(Renty & Roeyers, 2006)和降低社会成本(Lecciso et al., 2013)。

由于 2013 年美国精神病学会修订了《精神障碍诊断及统计手册》第五版(the Diagnostic and Statistical Manual of Mental Disorders-5th Edition, DSM-5), 并采用单一分类概念来定义自闭症谱系障碍，这一自闭症临床诊断标准上的巨大变化不仅对自闭症的发病率、诊断、治疗及预后其他相关领域产生深远影响(陈文雄, 2013), 更对确定 ASD 筛查行为特征提出了极大的挑战。因此，对现有的自闭症风险筛查工具进行质量分析显得十分重要和必要。

在过去的十几年里，已经有 19 篇英文文献综述对早期发现自闭症风险的筛查工具的评估标准进行了深入分析，我们将这些综述文献的优势和不足归纳为五大方面。在文献检索的全面性上，部分综述的结论仅搜索了 1~3 个数据库，甚至是完全基于经验性分析的(Charman & Gotham, 2013; Thabtah & Peebles, 2019); 在研究的地域分布上，19 篇综述侧重于欧洲或美国等特定地理区域(García-Primo et al., 2014; Levy et al., 2020), 来自亚非拉国家的研究相对少见(Soleimani et al., 2014); 在专业范畴上，则集中于医学和儿科学研究(Wang et al., 2020; Zwaigenbaum et al, 2015)及公共卫生领域(Stewart & Lee, 2017), 基于心理学和脑神经科学的专业指向较少涉及(Marlow et al., 2019)。在筛查工具从开发到实施所涉及的重要指标上，除了如上谈到的筛查工具要符合 DSM-5 的规范外，还有很多评估指标，如筛查样本的类型或应用水平(一级筛查或是二级筛查)、筛查信息来源(父母或看护者的报告，或是观察孩子的现场

(录像)互动)、抽样儿童的年龄和发育水平、筛查的实用性(如管理时间、工具的管理和评分、筛查人员的培训、转诊和随访结果管理等)、区分其他障碍的测量学特性(如敏感性和特异性、可鉴别性指标等)、心理测量学特性(重测信度和评分者信度、效度等)、文化适应性、目标受众的可理解性和欢迎程度、网络和移动设备的可访问性等 10 余项，但大多数综述只选择了 2~3 个指标，开展全部指标的评估较为困难。最后，在综述分析所采用的研究方法上，除 Hampton 和 Strand (2015)、Sánchez-García 等(2019)、Yuen 等(2018)等三个研究小组采用元分析的方法或应用贝叶斯层次模型来评估与准确性、心理测量学特性相关的结果外，大多数综述没有全面分析筛查工具的心理测量学和测量学特性。

对自闭症风险信号的早期筛查是一项极为艰巨的任务，来自不同学科背景(如遗传学、神经科学、精神病学、心理学、病毒学)的自闭症研究人员的共同目标在于确定能够在婴儿期和幼儿期可靠检测到的行为标志物，并将其纳入筛查测量中，但是不同的衡量标准对筛查工具的质量评价有不同的界定，不同筛查工具在目标行为的操作性和获取信息的方式上也有所不同。虽然以往的综述为我们理解筛查工具的功能和内容评价提供了良好的基础，但也显露出其在质量评价完整性上的不足。本研究对发表于 2013~2020 年间的适合学前阶段儿童自闭症谱系筛查工具进行了批判性评估，以识别其优点、性能问题和缺点。具体的研究问题包括：(1)从应用水平、心理测量学特性、文化适应性、符合诊断标准等多指标切入，详细地介绍了这些自闭症谱系筛查工具的特征表现；(2)通过综合比较婴儿和幼儿年龄阶段筛查工具在敏感性和特异性数据，探讨是否有一个(或多个)有希望用于早期筛查自闭症风险的工具？

2 研究方法

2.1 文献搜索策略

采用基于系统评价和元分析(PRISMA)指南来确定需要纳入研究的论文构成(Moher et al., 2009; Siddaway et al, 2019)，应用的英文主题词是: ‘ASD screen*’, ‘ASD detect*’, ‘ASD or autism or autist*’, ‘screening tool’, ‘Early identification’, ‘assessment tool’, ‘assess*’, ‘instrument*’, ‘measure*’, ‘tool*’，以“2013 年 1 月~2020 年 12 月之内”和“English 英语”作为高级过滤词。分别从 3 个综合性文摘检索数据库(WOS、SCOPUS、PubMed 等)检索到相应文献，通过链接进入 APA/EBSCO、ScienceDirect、Sage、Springlink、Wiley Online Library 电子期刊数据库，获取 5535 篇文献，如图 1 所示。电子搜索是由第二和第三作者进行的，他

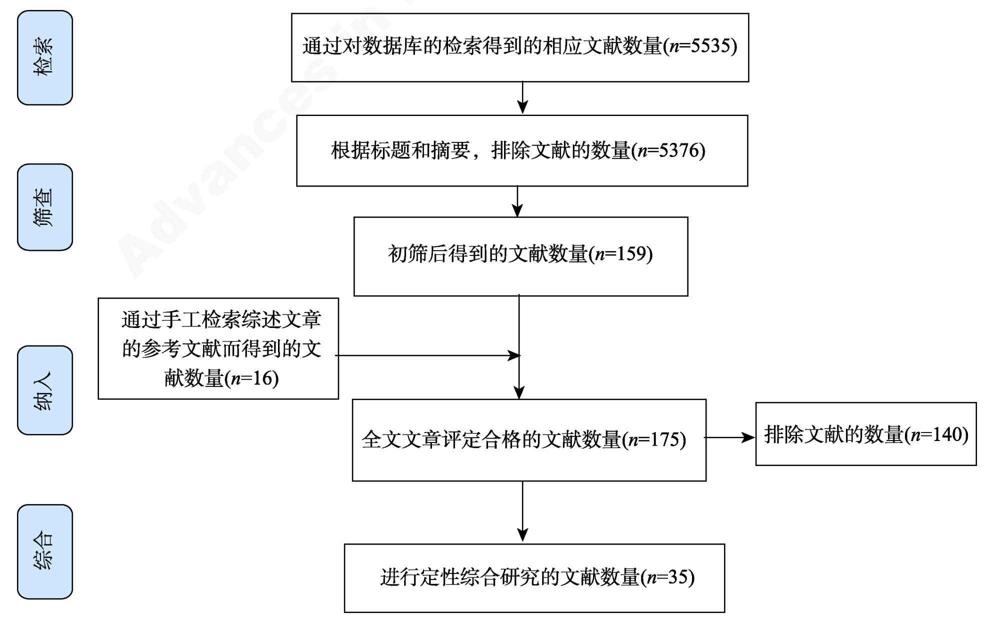


图 1 纳入 ASD 筛查工具综述的 PRISMA 流程图

们提取了这些文献记录, 并将参考资料制成表格, 放在 excel 文件中。第一作者依据长期从事自闭症认知研究和教学实践经验, 预先制定了文献纳入/排除标准, 所有作者共同仔细阅读这些引文的标题和摘要后, 删除重复、不合格论文, 获取初筛后文献 159 篇。此外, 从已经发表的"自闭症谱系筛查"综论文章的参考文献和引文中进行二次手工搜索, 也对同一主题的其他评论进行了检查, 提取符合我们研究的原创性研究文献 16 篇, 共计纳入质量评估的文献 175 篇。之后, 根据纳入/排除标准, 删除文献 120 篇, 根据质量评估标准的偏倚风险的高低, 删除高风险文献 20 篇, 最后纳入全文定性综合分析的文献 35 篇。

同时, 我们也应用中文主题词: '自闭症谱系筛查', '筛查 and/or 筛选', '筛查工具 and/or 筛选工具', '自闭症 and/or 孤独症'等, 获取中文文章 37 篇。但是, 根据文献纳入标准, 排除 23 篇文章：(1)综述性文献(n = 2); (2)年龄范围(0~6 岁)以外的样本群体(n = 4); (3) 不是以筛查为目的研究(n = 17), 如, 干预效果的说明, 测量工具信度效度的比较, 筛查模式和管理的评价, 对比其他发育障碍的流行病学等; 根据质量评估标准的偏倚风险的高低, 排除偏倚风险较高风险 14 篇文章。这些中文文章的偏倚风险主要体现在诊断标准和文化适应性上, 如在筛查的结果需要有明确的 ASD 诊断上, 虽然基本符合了"基于国际疾病分类(ICD)和/或美国精神障碍诊断和统计手册(DSM)的框架", 但在专家临床判断上, 没有研究应用了可靠和标准化的金标准测量, ADI-R 或 ADOS-2; 也没有关于评估员或诊断者的资格或培训信息的报告, 如陈述为"以临床医生根据 DSM-Ⅳ作出的临床诊断作为金标准" (龚郁杏 等, 2015); 同时, 大部分研究因未提供足够的分类准确性数据(骆名进 等, 2020), 无法确定用于识别自闭症儿童的筛查工具的有效性。其次, 在文化适应和管理上, 虽然少部分研究报告了翻译、回译的文化适应性调整, 但对量表项目数量的改变和临界取值的改变, 造成对这些工具的使用和分析更加复杂, 使用不同的版本或评分算法也导致我们无法标准化地提取相应数据, 如关智勇等(2016)年采用 CAST 中文修订版(37 项)发现, 对 3~< 4 岁儿童孤独症 CAST 临界值以 17 分为宜, 而由英国剑桥大学 Baron Cohen 团队开发研制的原版 CAST (Childrenhood Autism Spectrum Test)建议将临界值设为 15 分(Scott et al., 2002), 而许丹等(2013)的 CAST 中文修订版项目数量是 25 项, 少于英文原文 37 项, 但结果建议将 16 分作为 CAST 中文版的临界值。由此, 我们认为, 可以考虑单独设计基于中文相关研究成果的质量分析标准, 开展专项研究, 不适合纳入本文并与英文同类文献和工具进行横向比较。因此, 在本综述中, 并没有开展基于中文筛查工具的深度分析。

2.2　纳入和排除标准

正式确定的纳入标准包括: (1)在第一次电子搜索中确定的"自闭症谱系障碍筛查工具"是重点; (2)研究发表为"全文原创文章" (3)在同行评审期刊上发表的论文; (4)研究样本重叠的年龄范围为 0~6 岁; (5)研究样本是被监测 ASD 症状的个体, 即使他们有另一个主要诊断(例如, 如果探索筛查工具的测量特性, 监测"脆性 X 人群"表现出 ASD 症状的论文也是合格的); (6)研究目的是开发一种测量工具或评估其一种或多种测量特性的论文, 研究结果强调使用筛查工具评估 ASD 的准确度相关, 则将其纳入; (7)样本量大于等于 30 人。

排除标准包括: (1)只以"诊断"为目的研究的文章、对行为改变、进展或干预反应效果进行测量的文章(n = 24); (2)以区分发育障碍或其他类型障碍(如智力障碍或 ADHD 等), 而不是筛查 ASD 为目的的研究(n = 23); (3)针对 ASD 的流行病学研究和专家指南, 生物标志物、功能磁共振成像、血液检测、基因表达谱、共病、工具的机器学习算法等非本研究能力所及的检测程序(n = 27); (4)综述, 文摘, 报道, 会议论文等回顾性研究, 非原创性研究文献(n = 21); (5)年龄范围(0~6 岁)以外的样本群体, 无法区分年龄阶段的(n = 20); (6)非英文(n = 2); (7)无法获取全文(n = 3)。

2.3　质量评估的标准和评分者信度

QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies-2, 诊断准确性研究的质量评估-2)是广泛用于行为、医学领域(McDonald et al., 2015; Stout et al., 2015), 针对诊断和筛查测量文献开展质量评价的工具(Whiting et al., 2011), 包括诊断准确性研究中观察到的最重要的偏差和变异来源。在本研究中, 遵循 QUADAS-2 的循证程序, 完成 175 篇英文文献的质量标准评估。首先, 对给定的关于筛查工具的调查问题进行了调整:

(1)将被试样本量、符合参考标准(诊断金标准)、流程和时间、心理测量学指标(信度和效度)和分类准确性指标(敏感性和特异性等)作为偏倚风险维度审查维度; (2)为每个领域开发了探究性问题, 并分配了是/否评分。第二, 对问题的实用性进行实地测试, 即由三位作者从 175 篇文章中随机选定 3 篇, 背对背仔细阅读全文, 并将探究性问题应用于每个研究, 在不同的excel文件中报告了他们的决定, 并对他们的发现进行了逐条记录的比较。如有异议, 由第三提交人仲裁, 最后统一大家的共识。

3 结果

3.1 纳入研究的方法学质量分析

我们使用 QUADAS-2 质量评估研究工具和评分者信度系数来检验纳入研究总体质量分数和评分者之间的一致性, 并通过讨论解决任何项目差异。由第一作者另外培训两名不了解研究目的的研究生作为独立评分者, 帮助每名评分者准确了解 QUADAS-2 所涉及审查维度的定义, 明确对应探究性问题的范围, 能准确掌握评分的技巧和方法, 与前三位作者共同完成 175 篇文章的全文质量评估。根据质量评估标准的偏倚风险维度: 被试样本量、诊断标准、分类准确性、心理测量指标, 删除高风险文献 20 篇。最后, 随机选择 175 篇文献中的 50 篇, 采用徐建平和张厚粲(2005)提出的"归类一致性公式"计算出评分者信度在 0.64~0.94 之间, 总的归类一致性为 0.75; 编码信度系数值在 0.78~0.97 之间, 总体编码信度系数为 0.86, 这表明了本次质量评估的评分者信度具有公平到良好的一致性水平。

3.2 适合学前儿童自闭症风险预警特征的筛查工具概述

为了全面清晰地掌握学前儿童自闭风险筛查工具的基本特性, 我们分别报告了筛查工具的名称、作者国别和研究年份、样本特征、应用水平、心理测量学特性、文化适应性和诊断标准、筛查分类的准确性等情况。

我们从质量评估后的 35 篇文献中提取出筛查工具 18 项, 表 1 和表 2 分别报告了每项筛查工具的基本内容和发表文献的一般细节。在 2013 年~2020 年的 7 年间, 文献发表时间在 2019 年最多, 9 篇, 其次是 2014 年 6 篇, 2018 年 5 篇, 平均每年发表文献数量是 4 篇。在地点上, 大多数研究(n = 16)是在美洲进行的, 12 项在美国。来自亚洲国家的研究有 12 项, 5 项在中国, 欧洲的研究有 7 项, 意大利有 2 项。按照筛查儿童样本的年龄范围划分为两个年龄阶段, 有 21 项研究中使用了 10 项筛查工具探讨了 0~3 岁婴儿的自闭症风险, 样本平均数量 4047 人, 性别比例男童大于女童(3101:1937[1]); 样本取样的平均年龄范围在 15.62~26.94 个月(M = 20.62)[2]; 在 3~6 岁的幼儿阶段, 有 18 项研究使用了 14 项工具[3]分析了自闭症风险特征, 参与儿童的平均数量是 606 人, 性别比例男童大于女童(371:235); 样本的平均年龄范围在 22.36~51.97 个月(M = 37.26), 见表 3、表 4。

3.2.1 筛查工具在应用水平上的表现

自 20 世纪 90 年代末开始专门针对自闭症进行筛查工具的研制, 主要是根据筛查样本的类型划分为一级和二级筛查两种应用水平。一级筛查(Level I)适用于未经选择的一般人群儿童, 无论其风险状况如何。二级筛查(Level II)针对的是已经被确定为发育障碍儿童的筛查(例如, 有自闭症谱系家族史、父母或临床医生提出了担忧、经过一级筛查鉴定为阳性者)。但我们在梳理文献的过程中发现, 实践中对测量工具的应用水平并没有那么明晰的区分界限, 借鉴 Robins 和 Dumont-Mathieu (2006)将 M-CHAT 作为同时应用于一级和二级筛查的介绍, 我们增加了"混合(hybrid)"工具的应用水平, 即指那些同时应用于一般人群(一级筛查)和临床样本(二级筛查)的筛查工具, 详见表 2。

在本研究中, 我们发现了 4 个一级筛查工具, 即: 改良的幼儿自闭症检查表(M-CHAT); 第一年调查(FYI); 心理发展问卷-1 (PDQ-1); 自闭症谱

---

[1] 作者注: 4 项研究性别有缺失值, 因此, 男童与女童合计平均值与总样本平均值相差 13 人, 详见表 3。

[2] 作者注: 在婴儿阶段, 平均抽样年龄的最小值是 16 个月, 只有 2 项研究的实际抽样年龄在 16 个月前(Sacrey et al, 2018; Turner-Brown et al, 2013), 但这 2 个研究在筛查工具和取样年龄均不相同, 因此, 无法详细介绍指向 16 个月前筛查工具。

[3] 作者注: 有 4 篇文献采用的是纵向追踪设计 M-CHAT (Toh et al, 2018; Koh et al, 2014)、M-CHAT-R/F (Magán-Maganto et al, 2020)、BISCUIT-Part 1 (Horovitz & Matson, 2014)同时测量了婴儿和幼儿两组, 所以同时出现在表 3 和表 4 中。

表 1 纳入本研究中自闭症风险筛查工具的详细描述

测量名称(简称)	测量的结构	适用年龄	项目数	计分	时间分钟	可获得性	测量方式	临界值	研究数量
1. 改良的幼儿自闭症检查表(the Modified Checklist for Autism in Toddlers, M-CHAT)	7 个领域: 社会利益, 假装功能性游戏, 共同注意, 原陈述式指点, 模仿, 运动发展, 吵闹式的玩耍	16~30 个月	23	是/否	5~10	免费	主要照顾者完成问卷	对于 6 个关键项目中 2 分及以上答案为否; 或者 3 分及以上条目答案为否时, 为阳性结果, 怀疑为自闭症患儿	4 项研究 Carbone 等(2020) Toh 等(2018) Stenberg 等(2014) Koh 等(2014)
2. 改良的幼儿自闭症检查表−有后续随访(the Modified Checklist for Autism in Toddlers with Follow-up Interview, M-CHAT/F)	7 个领域: 社会兴趣, 假装功能性游戏, 共同注意, 原陈述式指点, 模仿, 运动发展, 吵闹式的玩耍	16~30 个月	23	是/否	10~15	免费	主要照顾者完成问卷, 如果孩子的筛查结果呈阳性, 附加专业人员随访	对于 6 个关键项目中 2 分及以上答案为否; 或者 3 分及以上条目答案为否时, 为阳性结果, 怀疑为自闭症患儿	6 项研究 Guthrie 等(2019) Kerub 等(2020) Baduel 等(2017) Sturner 等(2016) Srisinghasongkram 等(2016) Kamio 等(2014)
3. 改良的幼儿自闭症检查表−有修订的后续随访(Modified Checklist for Autism in Toddlers–Revised with follow-up, M-CHAT-R/F)	7 个领域: 社会兴趣, 假装功能性游戏, 共同注意, 原陈述式指点, 模仿, 运动发展, 吵闹式的玩耍	16~30 个月	20	通过/不通过	10~20	免费	主要照顾者完成问卷, 如果孩子的筛查结果大于 3 分, 专业人员开展结构化的后续问题,	0~2 分: 低风险, 无需进一步评估, 除非存在其他风险因素 3~7 分: 中等风险, 需要进行随访以确定是否需要转诊 8~20 分: 高风险, 立即转诊进行评估和干预	5 项研究 Coelho-Medeiros 等(2019) Magán-Maganto 等(2020) Guo 等(2019) Robins 等(2014) Tsai 等(2019)
4. 幼儿自闭症定量检查表(Qualitative Checklist for Autism in Toddlers, Q-CHAT)	3 个部分: 社会交往; 行为; 语言	18~24 个月	25	5 点利克特量表	15~20	免费	主要照顾者完成问卷	Q-CHAT: 27 分 Q-CHAT-10: 3 分	2 项研究 Raza 等(2019) Ruta 等(2019)
5. 儿童行为检查表(the Child Behavior Checklist, CBCL)	7 个症状量表(情绪反应、焦虑/抑郁、躯体抱怨、退缩、睡眠问题、注意力问题和攻击行为)和 5 个 DSM 导向的量表(情感问题、焦虑问题、PDP、注意力缺陷/多动问题和对立违抗问题)	学龄前版本(1.5~5 岁)	100	3 点利克特量表	15~20	免费	主要照顾者完成问卷	临床意义: 症状量表和 DSM 导向量表的 T 分数为 65~69 分以上; 内化、外化和总问题量表为 60~63 分以上	4 项研究 Rescorla 等(2019) Limberg 等(2017) Narzisi 等(2013) Havdahl 等(2016)

续表

测量名称(简称)	测量的结构	适用年龄	项目数	计分	时间分钟	可获得性	测量方式	临界值	研究数量
6. 婴幼儿自闭症筛查第 1 部分(The Baby and Infant Screen for Children with a UtIsm Traits—Part1, BISCUIT-Part1)	3 个因素：社会化/非语言交流、重复行为/兴趣和交流受限	17~37 个月	62	3 点利克特量表	15	免费	主要照顾者完成问卷	17 分以下：ASD/非典型发展 17~39 分：ASD/PDD-NOS 39 分或以上：ASD/autism	1 项研究 Horovitz 和 Matson (2014)
7. 婴儿自闭症父母筛查(The Autism Parent Screen for Infants, APSI)	包括眼神接触、视觉跟踪、对姓名的反应、模仿、语言、社会发展、共同注意、手势、游戏、物体视觉检查和情绪调节的损伤	6~24 个月	26	3 点利克特量表	10~15	免费	主要照顾者完成问卷	15 分	1 项研究 Sacrey 等(2018)
8. 第一年调查(The First Year Inventory, FYI)	2 个领域：社会沟通和感官调节	12 个月	63	3 点利克特量表	10~15	免费	主要照顾者完成问卷	总风险得分：19.2 分(等于或高于 96 百分位) 社会沟通领域：22.5 分(第 94 个百分位) 感觉-调节领域：14.75 分(第 88 百分位)	1 项研究 Turner-Brown 等(2013)
9. 心理发展问卷-1(the Psychological Development Questionnaire-1, PDQ-1)	2 个领域：社会参照和交流表达	18~36 个月	10	3 点利克特量表	2	免费	主要照顾者完成问卷	12 分	1 项研究 Zahorodny 等(2018)
10. 自闭症谱系评定量表中文修订版(the revised Chinese version of the autism spectrum rating scale, RC_ASRS)	2 大因素：社交、异常行为	2~5 岁	62	4 点利克特量表	20	付费	主要照顾者完成问卷	60 分	1 项研究 Zhou 等(2018)
11. 自闭症图片评估清单(the pictorial autism assessment schedule, PAAS)	2 个领域：异常的游戏行为和社交互动困难。清单的前 20 项都配了一张照片，用文字说明了信息，21 项是文本信息	18~48 个月	21	是否计分	15~20	免费	主要照顾者完成评估清单	4 分	1 项研究 Perera 等(2017)
12. 发展性登记(The Developmental Check-In, DCI)	4 个领域：交流、游戏、社交和行为。28 个项目每一个都被描绘成一幅线条图和一张照片。添加了简短的描述标签。	24~60 个月	28	是否计分	NR	NR	主要照顾者完成的问卷	7.5 分	1 项研究 Janvier 等(2019)

续表

测量名称(简称)	测量的结构	适用年龄	项目数	计分	时间分钟	可获得性	测量方式	临界值	研究数量
13. 两岁儿童自闭症筛查测验(Screening Tool for Autism in 2-Year-Olds, STAT)	4 个社会交际领域: 游戏, 请求、共同注意、模仿	24~35 个月	12	通过或失败	20	免费	专业人员对儿童进行游戏互动观察、判断	0 到 4 分之间分数越高, 障碍越严重。 24 个月以下儿童: 2.5 分 24 个月以上儿童: 2.75 分	2 项研究 Wu 等(2020) Chiang 等(2013)
14. 儿童早期自闭症检测(the Autism Detection in Early Childhood, ADEC)	5 个部分: 社会互动、言语和非言语交流、活动或兴趣受限、对感官刺激固着性反应	12~36 个月	16	3 点利克特量表	10~15	NR	专业人员对儿童进行观察、判断	0~10 分: ASD 低风险 11~13: ASD 中风险 14~19: ASD 高风险 大于 19 分: 非常高风险	1 项研究 Hedley 等(2015)
15. 自闭症心理状态检查(the Autism Mental Status Exam, AMSE)	4 个方面: 社会, 沟通, 自闭症的行为标记和症状	3~6 岁	8	3 点利克特量表	NR	NR	专业人员基于儿童互动的观察和照顾者访谈	7.5 分	1 项研究 Cederlund (2019)
16. 三项直接观察的自闭症筛查(The Three-Item Direct Observation Screen, TIDOS)	3 个项目: 叫名反应; 共同注意; 眼神接触	18~60 个月	3	前两项 0~2 分, 后一项 0~1 计分	15~20	免费	专业人员的观测	1 分	1 项研究 Oner 等(2014)
17. 幼儿自闭症快速互动筛查测试(the Rapid Interactive Screening Test for Autism in Toddlers, RITA-T)	5 个方面: 共同注意、社会意识、情绪反应、人类能动性意识和一些基本认知技能。	18~36 个月	9 个活动项目最高 30 分	是/否计分	10	免费	专业人员对互动活动“触发儿童的特定行为进行观察	15 分	1 项研究 Choueiri 和 Wagner (2015)
18. 自闭症追踪的结构化观察(Observação estruturada para rastreamento de Autism, OERA) [葡萄牙语]	2 个领域: 缺乏社交技能和限制性、重复性的行为或兴趣	3~10 岁	13	是否计分	10~15	免费	专业人员使用一套标准化的物品和玩具与孩子玩耍并观察	5 分	1 项研究 Paula 等(2018)

表 2　18 项自闭症风险筛查工具在应用水平、心理测量、文化适应性和诊断标准方面的特征

工具名称	研究者,年代	应用水平	心理测量指标		文化适应性		诊断标准		
			信度	效度	转译和回译	审查与预测试	发育及其他相关测量	诊断工具	诊断标准
M-CHAT (23 项)	Carbone 等(2020)美国	一级	NR	NR	NR	√	NR	NR	ICD-9 ICD-10
	Toh 等(2018)马来西亚	一级	NR	NR	√√	√	NR	NR	DSM-IV-TR
	Stenberg 等(2014)挪威	一级	NR	NR	√√	√√	NR	ADI-R ADOS	ICD-10 DSM-IV
	Koh 等(2014)新加坡	一级	α: 0.75~0.86	NR	√√	√	Vineland-II Bayley-III WPPSI-III DAS-II	ADOS-G	DSM-IV-TR
M-CHAT/F (23 项)	Guthrie 等(2019)美国	一级	NR	NR	–	–	NR	NR	ICD-9 ICD-10 DSM-IV DSM-5
	Kerub 等(2020)以色列	二级	NR	NR	√√	NR	GDS	NR	DSM-5
	Baduel 等(2017)法国	二级	NR	NR	√√	NR	PEP-R VABS	ADOS-G	NR
	Sturner 等(2016)美国	一级	评分者信度 $r = 0.866$ Cohen's K= 0.72	NR	–	–	MSEL	ADOS-2	DSM-IV
	Srisinghasongkram 等(2016)泰国	混合	NR	NR	√√	NR	MSEL CAT/CLAMS 或 Denver II	NR	DSM-5
	Kamio 等(2014)日本	混合	NR	NR	√√	√	TBIS EADT WISC-III	ADI-R ADOS CARS	DSM-IV-TR
M-CHAT-R/F (20 项)	Coelho-Medeiros 等(2019)智利	混合	α = 0.889	与 ADOS-2 相关 $r = 0.849$	√√	NR	NR	ADOS-2	DSM-5
	Magán-Maganto 等(2020)西班牙	一级	α: 0.62~0.72	NR	√√	√√	M-PRS VABS	ADOS-G ADOS-2	DSM-5
	Guo 等(2019)中国	一级	α = 0.86 重测信度 ICC = 0.759	NR	√√	√	ASQ-3 ASQ-SE CARS	CARS	DSM-5
	Robins 等(2014)美国	一级	α = 0.79	NR	–	–	CARS-2 TASI MSEL VABS-II BASC-2 STAT	ADOS	DSM-IV-TR
	Tsai 等(2019)中国台湾	混合	α = 0.84 重测信度 K = 0.33~1 ICC = 0.94~0.96	与 CBCL/1½-5 相关 $r = 0.37$~0.63	√√	√√	CBCL/1½-5 MSEL	NR	DSM-5
PDQ-1	Zahorodny 等(2018)美国	一级	重测信度 $r = 0.997$	与 ABC 相关 $r = -0.869$	–	–	ABC MSEL	ADI-R	DSM-IV-TR

续表

工具名称	研究者, 年代	应用水平	心理测量指标		文化适应性		诊断标准		
			信度	效度	转译和回译	审查与预测试	发育及其他相关测量	诊断工具	诊断标准
FYI	Turner-Brown 等(2013)美国	一级	NR	NR	–	–	SRS-P DCQ MSEL VABS-2	ADOS	DSM-IV
RC_ASRS	Zhou 等(2018)中国	一级	α = 0.91	Cohen's d = 2.27	√√	√√	NR	NR	DSM-5
Q-CHAT	Raza 等(2019)加拿大	二级	NR	NR	–	–	MSEL	ADOS ADI-R	DSM-IV
	Ruta 等(2019)意大利	混合	α = 0.84	与 CBCL/1½-5 相关 ρ 为 0.29–0.44	√√	√	GMDS CBCL/1½-5	ADOS-2	DSM-5
CBCL/1½-5	Rescorla 等(2019)美国	二级	退缩: α = 0.75 PDP: α = 0.80	NR	–	–	CDI MSEL	ADOS-G ADI-R	DSM-IV
	Limberg 等(2017)德国	一级	NR	NR	NR	NR	NR	ADOS-G ADI-R	ICD-10
	Narzisi 等(2013)意大利	混合	重测信度 r = 0.085 评分者一致性 r = 0.61	NR	NR	√	GMDS	ADOS-G CARS	DSM-IV-TR
	Havdahl 等(2016)美国	二级	NR	NR	–	–	DAS-II 或 MSEL CPRS-R SCAS MASC VABS-II	ADOS ADI-R	DSM-IV-TR
STAT	Wu 等(2020)中国台湾	二级	α = 0.9	NR	NR	NR	MSEL	ADOS	DSM-5
	Chiang 等(2013)中国台湾	二级	评分者信度 r = 0.90	NR	NR	√	MSEL	ADOS	DSM-IV DSM-IV-TR
APSI	Sacrey 等(2018)加拿大	二级	α: 0.77~0.92, 分半信度 r: 0.35~0.94	NR	–	–	AOSI MSEL VABS	ADOS ADI-R	DSM-IV-TR
BISCUIT-Part1	Horovitz 和 Matson (2014)美国	二级	NR	NR	–	–	BDI-2 M-CHAT	NR	DSM-IV-TR
ADEC	Hedley 等(2015)美国	二级	α = 0.8; 评分者信度 ICC = 0.95 K: 0.43~0.89	ADEC 总分与 CARS2-ST 和 ADOS-2 相关 0.58~0.70	–	–	CARS2-ST MSEL ASRS VinelandII BayleyIII	ADOS-2 ADI-R	DSM-5
AMSE	Cederlund (2019)瑞典	二级	NR	NR	–	–	VinelandII WPPSI 系列 M-PRS BSITD-3	DISCO-11	ICD-10
RITA-T	Choueiri 和 Wagner (2015)美国	二级	NR	与 ADOS-G 相关 r = 0.79	–	–	M-CHAT MSEL	ADOS-G	DSM-IV DSM-5

续表

工具名称	研究者, 年代	应用水平	心理测量指标		文化适应性		诊断标准		
			信度	效度	转译和回译	审查与预测试	发育及其他相关测量	诊断工具	诊断标准
DCI	Janvier 等(2019)美国	二级	NR	NR	–	–	M-CHAT-R SCQ MSEL	ADOS-2	NR
PAAS	Perera 等(2017)斯里兰卡	混合	重测信度 r = 0.957	NR	–	–	NR	CARS	DSM-5
TIDOS	Oner 等(2014)土耳其	混合	评分者一致性 0.80 及以上	NR	NR	NR	SCQ VABS	ADOS	DSM-IV DSM-IV-TR
OERA	Paula 等(2018)巴西	二级	评分者信度 K:0.652~0.978	NR	–	–	WISC-IV SON-R2½-7 ABC	ADI-R	DSM-5

注：–指该项研究不需要进行文化适应性检验, √指满足标题中的 1 项指标, NR = Not Report 没有报告

系评定量表中文修订版(RC_ASRS)。这些一级筛查的工具主要是对一般人群的广泛筛查和其他发育障碍的检查, 主要是基于父母报告的结果。但是, 由于一级筛查的阳性结果有可能导致 ASD 评估的过度参考、治疗的延迟、父母焦虑的增加以及稀缺资源的负担, 因此, 要对于筛查结果为阳性的样本推荐进入二级筛查, 旨在排除有其他类型发育障碍, 确认尚未做出鉴别诊断的儿童中的自闭症风险。本研究中有 8 个二级筛查工具, 包括：两岁儿童自闭症筛查测验(STAT)、婴儿自闭症父母筛查(APSI)、婴幼儿自闭症筛查第 1 部分(BISCUIT-Part1)、儿童早期自闭症检测(ADEC)、自闭症心理状态检查(AMSE)、自闭症追踪的结构化观察(OERA)、幼儿自闭症快速互动筛查测试(RITA-T)、发展性登记(DCI)。二级筛查主要依据训练有素的临床医生, 或是经过培训的专业人员直接与儿童互动并观察结果(Norris & Lecavalier, 2010)。与一级筛查不同的是, 二级筛查的对照组儿童可区分为两组不同的情况。第一组包括由于发育问题已经受到当地精神卫生服务机构关注的儿童、怀疑发育迟缓(DD)、语言障碍(LD)的儿童, 或符合非自闭症谱系(non-ASD)标准的儿童, 如脑瘫、癫痫、唐氏综合征、失明、哮喘、过敏的儿童, 这一群体被确定为发展关注组(DC), 也称为低危组(LR)。第二组包括患有自闭症儿童的兄弟姐妹, 或被一级筛查为自闭症阳性的儿童作为对照组, 被定义为遗传风险组(GR), 因为他们患自闭症的概率很高(Grønborgetal et al., 2013), 也称为高危组(HR)。纳入二级筛查的研究旨在: (a)通过 LR 或 HR 组来评价筛查工具的分类准确性; (b)比较 LR 和 HR 组之间的差异; (c)跟踪 LR/HR 组直到确诊, 通过减少假阳性的数量来减少专门 ASD 诊断评估的长等待名单; 或者, 最后, (d)将普通人群中的儿童与 LR 或 HR 组进行比较。

根据研究所选用样本的类型, 本研究确定了 6 项"混合"应用的测量工具, 即：改良的幼儿自闭症检查表–有后续随访(M-CHAT/F); 改良的幼儿自闭症检查表–有修订的后续随访(M-CHAT-R/F); 幼儿自闭症定量检查表(Q-CHAT); 儿童行为检查表学前儿童版(CBCL/1½-5); 三项直接观察的自闭症筛查(TIDOS); 自闭症图片评估清单(PAAS)。这些混合水平的筛查旨在检验: (a)将一般人群筛查工具应用于临床样本的适当性, (b)比较不同来源样本在筛查工具上的差异(ASD 与 ATD 与 ODD), 最后, (c)比较一般人群中的儿童(TD)与 ASD 诊断儿童的差异。

3.2.2 筛查工具在心理测量特性上的表现

根据 Cicchetti (1994)的指南, 分析了筛查工具的内部一致性、稳定性、评分者间信度和重测信度。他们指出, 低于克伦巴赫 α 系数 0.70 意味着临床意义的水平是不可接受的, 0.70 到 0.79 表示临床意义水平为一般, 0.80 到 0.89 良好, 高于 0.90 表示临床意义水平为极好。Cohen 的 kappa、加权 kappa 和类内相关(ICC)被用来衡量评分者之间的一致性和重测信度。Person 的积矩相关(r)测量了独立评分者在测试分数排序上的相似性, 它

也被用来测量一些研究的可靠性。Kappa 或 ICC 统计值低于 0.40 表示临床意义较差，介于 0.40 和 0.59 之间表示临床意义一般, 0.60~ 0.74 表示良好, 0.75 及以上表示具有非常好的临床意义(Cicchetti et al., 1995)。

本研究中筛查工具的信度检验结果表明，有 12 项研究(占 35 项研究中的 34%)使用克伦巴赫 α 系数检验信度，但 α 值在 0.80 以上的工具只有 6 种: RC_ASRS、STAT、M-CHAT-R/F、Q-CHAT、ADEC、CBCL/1½-5 的 PDP 量表。评分者信度采用 ICC 或相关系数衡量，且信度值在 0.75 以上的工具有 5 种：M-CHAT-R/F、STAT、ADEC、TIDOS、M-CHAT/F; Kappa 值检验评分者信度的有 4 种工具，最高是 Paula 等(2018)研究所采用的 OERA (0.652~0.978)，最低是 M-CHAT/F (0.42) (Kerub et al., 2020)。使用重测信度检验的工具有 5 种，PDQ-1 的重测信度值最高，r = 0.997 (Zahorodny et al., 2018); CBCL/1½-5 量表的重测信度最低 r = 0.085，时间间隔是 8 天(Narzisi et al., 2013)。

由于效度的高低没有明确的标准，不同效度标准也不一样，我们通常的最低标准：效度系数必须在统计上有显著性，即此相关系数不是来自相关系数的参数 ρ 为零的全域(魏华忠, 1995)。在本研究纳入的 35 项研究中，有 7 项(占 20%)研究检验了筛查工具的效标效度，相关系数达到显著水平，这也是最多使用的效度检验方式, M-CHAT-R/F, ADEC, RITA-T 与诊断工具(ADOS\ CARS)相关在 0.58~0.849 之间; PDQ-1, OERA 与效标 ABC 相关 0.4~0.869，M-CHAT-R/F 与效标 CBCL 的 Pearson 相关在 0.37~0.63 之间, Q-CHAT 与效标 CBCL 的 spearman ρ 相关在 0.29~0.44 之间。有 15 项(占 43%)没有进行信度和效度检验。

3.2.3　筛查工具在文化适应性的表现

不同的国家和文化背景会对人类行为期望有着巨大的影响，不同文化下所界定的行为也表现出巨大的异质性，ASD 筛查工具开发和行为诊断也因此受到极大的挑战(Wallis & Pinto-Martin, 2008)。因此，本研究将筛查工具的文化适应性作为一个重要指标进行检查，遵循既定的语言和文化等效性准则，分别从翻译、回译、委员会审查和实地预测试这四方面，对每项研究的文化适应性进行独立评估(Guillemin et al., 1993; Soto et al., 2015)。排除 13 项由英语国家开发并以英语作为测量语言的研究, 3 项最初是使用非英语的语言开发的也应用于当地人群样本的研究。最后，有 19 项研究需要进行文化适应性的检验，但其中只有 4 项研究符合完整准则(Stenberg et al., 2014; Magán-Maganto et al., 2020; Tsai et al., 2019; Zhou et al., 2018)。这 4 项研究不仅说明了筛查项目在翻译、回译上进行措辞和语言上的调整，项目也经过伦理委员会的审查，还详细地报告了预测试的内容，如“招募了少量(30 组家庭以内)家长对工具的句子和措辞的适用性进行评估，确保没有任何误解等”。另有 6 项研究没有介绍对工具进行怎样的修改，既没有描述前向的翻译，也没有出现或没有描述后向的回译，其中 2 项研究只介绍了是经过机构委员会审查的, 1 项研究只提到进行一次预测试，但没有描述预测试的内容和过程，见表 2。

值得强调的是，更广泛的文化适应性研究表明，需要以受访者能够理解或认为合适的方式来构建项目，如使用基于图片的插图，或使用当地儿童的照片等。例如，斯里兰卡的 Perera (2017)设计了一种基于文化适应的自闭症谱系障碍图形筛查工具，自闭症图片评估清单(PAAS)，将陈述 ASD 主要特征的每个项目与一张照片配对，以提高理解能力。这项新工具在自闭症儿童、非自闭症发育障碍儿童和普通儿童样本中进行了测试，表现出很高的敏感性和特异性(Se = 0.88, Sp = 0.933)。Janvier 等(2019)针对服务不足社区(低收入、少数民族家庭或英语水平有限的家庭)幼儿开发的自闭症筛查工具，发展性登记(DCI)，也是用图片来说明目标行为的一种新的工具，在交流、游戏、社交和行为 4 个领域中设计了 28 个项目，每一个项目都被描绘成一幅线条图和一张照片。在 24~60 个月之间的儿童样本中进行验证的结果表明，它能很好地区分自闭症谱系障碍和非自闭症谱系障碍(Se = 0.66, Sp = 0.76)。

3.2.4　筛查工具符合诊断标准的表现

虽然 DSM-5 自 2013 年颁布使用之后得到广泛地应用，在 35 篇筛查测量文章中有 12 篇研究是符合单独的 DSM-5 标准的，占 34%; 而单独和/或使用 DSM-IV、DSM-IV-TR 的筛查研究有 15 篇，占 43%，仍然占比最大，这表明大多数筛查工具都遵循的是第四版或修订版本的精神障碍诊断和统计手册，而不是现行 DSM-5 手册的程序和

指南。单独参考 ICD-9 和 ICD-10 标准的只有 3 篇, 占 8%; 同时参考国际疾病分类(ICD)和美国精神障碍诊断和统计手册(DSM)有 2 项研究, 且都是样本超过 1 万以上普通儿童参与的一级筛查。诊断工具同时使用自闭症诊断观察表(ADOS)和自闭症诊断访谈-修订版(ADI-R)的研究有 8 项, 单独使用 ADOS 有 14 项, 单独使用 ADI-R 有 2 项。没有报告诊断工具的有 8 项, 见表 2。

在从筛查到诊断这段时间内, 绝大多数研究同时收集了样本的发育特征、智商、语言能力与情绪状态的数据, 目的是将之与临床判断结合使用, 以做出更有效的 ASD 诊断(Le Couteur et al., 2008)。最多使用的发育测量工具是穆伦早期学习量表(MSEL), 有 16 项; 采用文兰适应性行为量表的研究有 9 项; 智力测验有 6 项; 其他发展量表如, 贝利婴幼儿发展量表(BSITD-3)有 3 项; 美林−帕尔默修订的发展量表(M－PRS)、巴特尔发展量表第 2 版(BDI-2)等共计 9 项; 没有报告发育相关测量的研究有 8 项。

3.3 不同年龄阶段筛查工具的分类准确性分析

根据心理学关于儿童心理发展年龄阶段的划分标准, 我们将抽样年龄在 0~36 个月(3 岁)的研究纳入婴儿阶段, 而 36~72 个月(6 岁)的研究纳入幼儿阶段(杨清, 1985), 然后详细地分析每一种筛查工具的分类准确性, 用以发现不同年龄阶段筛查工具的辨别力以及产生误导性结论的可能性。本研究主要评估了每种筛查工具的 4 个重要参数: (1)敏感度(Sensitivity)是指筛查结果为阳性的患者比例, 是指 ASD 儿童根据筛查结果被正确识别为“高危”的比例; 自闭症谱系障碍的儿童如果没有在筛查中被识别出来, 就会被认为是假阴性(False-negative); (2)特异性(Specificity)是指筛查结果为阴性的患者比例, 指没有自闭症谱系障碍的儿童使用筛查工具被正确分类为没有自闭症谱系障碍风险的比例; 没有自闭症谱系障碍的儿童, 如果筛查呈阳性, 则被认为是假阳性(False-positive); (3)阳性预测值(Positive predictive value, PPV)是指筛查结果为阳性的个体中患有该疾病的比例, 即 ASD 诊断为 ASD 的比例除以筛查呈阳性的总病例数。(4)阴性预测值(Negative predictive value, NPV)指筛查阴性的儿童未被诊断为 ASD 的比例。PPV 和 NPV 受正在筛查人群中 ASD 基线患病率以及筛查工具的敏感性和特异性的影响。虽然敏感性和特异性是检测性能的内在指标, 但 PPV 和 NPV 对个体家庭和系统水平的筛查评价可能具有更内在的意义。

借鉴 Falkmer 等(2013)的做法, 为了评估两个年龄阶段最有辨别力的自闭症风险筛查工具, 分别将每一种工具的敏感性(Se)和特异性(Sp)分别合并产生了一个正确分类平均值, 参考 Cicchetti 等(1995)提出的诊断准确性指标(敏感性、特异性、PPV 和 NPV)指南: 差 =< 0.70; 一般 = 0.70~0.79; 良好 = 0.80~0.89; 优秀 = 0.90~1.00。在婴儿阶段可以优先选取以下 3 种筛查工具: M-CHAT-R/F、STAT、PDQ-1。根据图 2 可知, 这 3 种筛查工具的诊断准确性指标达到了良好水平, 即 Se 与 Sp 均大于 0.8, 因此, 相较于其他的几种筛查工具能够提供更准确的筛查结果。而在幼儿阶段, OERA、TIDOS 筛查工具的诊断准确性达到优秀水平, 即 Se 与 Sp 均大于 0.9, 在对该年龄阶段的被试筛查时, 可以优先选择这两种筛查工具,同时, 还有 6 种在幼儿阶段的筛查工具也达到了良好水平: RITA-T、T-STAT、RC_ASRS、PAAS、BISCUIT-Part1、M-CHATR/F, 见图 3。纵向比较

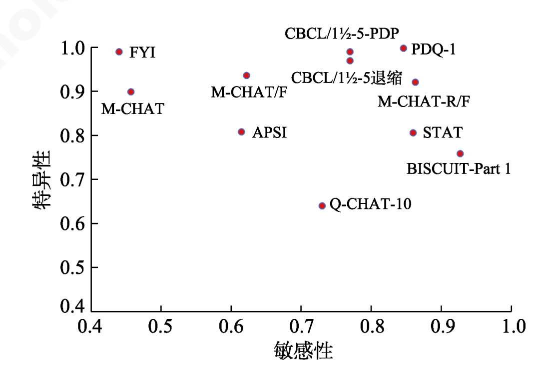


图 2 婴儿阶段自闭症风险筛查工具的敏感性和特异性

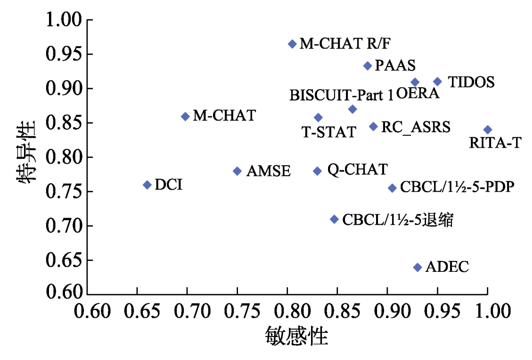


图 3 幼儿阶段自闭症风险筛查工具的敏感性和特异性

两个年龄阶段的分类准确性, 我们发现, 幼儿筛查工具的准确率的平均值(Se = 0.86, Sp = 0.83)要好于婴儿筛查工具(Se = 0.72, Sp = 0.88); 筛查的阳性率(PPV = 0.80)也要高于婴儿阶段(PPV = 0.59), 参见表 3、表 4。

4　讨论

我们着重从筛查工具的基本情况、应用水平、心理测量特性、文化适应性、符合诊断标准, 以及分类准确性方面评估了上述 18 项筛查工具, 这不仅有助于全面认识自闭症筛查工具、性能和可能出现的错误, 也为筛查工具的改进和创新提供了机会。

4.1　规范工具的质量评估标准, 开展持续的自闭症风险监测和评估

本综述期望回答的第一个问题是明确这些经过 QUADAS-2 质量评估后, 纳入筛查工具在应用水平、心理测量学特性、文化适应性和符合诊断标准等方面的表现。首先在筛查工具应用方面, 为什么幼儿阶段的筛查工具数量多, 准确性高,

表 3　婴儿阶段自闭症风险筛查工具的分类准确性

工具名称	研究者, 年代	样本数量/性别(男:女)/年龄(平均值)	敏感性	特异性	阳性预测值 PPV	阴性预测值 NPV
M-CHAT(23 项)	Carbone 等(2020) Koh 等(2014) Stenberg 等(2014) Toh 等(2018)	23966/12264:11702 (1 人性别不明)/17.25~26.25 mon	0.46	0.90	0.31	0.98
M-CHAT/F (23 项)	Baduel 等(2017) Guthrie 等(2019) Kamio 等(2014) Kerub 等(2020) Srisinghasongkram 等(2016) Sturner 等(2016)	5272/2735:2537/16.74~31.76 mon (22.18)	0.62	0.94	0.49	0.99
M-CHAT-R/F (20 项)	Coelho-Medeiros 等(2019) Guo 等(2019) Magán-Maganto 等(2020) Robins 等(2014)	6886/3582:3239(259 性别不明)/16~26.5 mon ($M = 21.08$)	0.86	0.92	0.35	1.00
PDQ-1	Zahorodny 等(2018)	1959/1009:950/18~36 mon	0.8462	0.9984	0.8800	0.9979
FYI	Turner-Brown 等(2013)	699/343:356/12 mon	0.44	0.99	0.31	0.99
STAT	Wu 等(2020)	139/91:28/16~24 mon (20 人未参加随访)	0.860	0.806	0.803	0.862
BISCUIT-Part1	Horovitz 和 Matson (2014)	1014/684:326(4 人性别不明)/17~23 mon ($M = 20.43$)	0.9274	0.7593	NR	NR
Q-CHAT-10	Raza 等(2019)	172/89:83/18~24 mon	0.73	0.64	0.35	0.90
CBCL/1½-5	Rescorla 等(2019)	82/52:30/24 mon　退缩	0.77	0.97	0.83	0.96
		PDP	0.77	0.99	0.91	0.96
APSI	Sacrey 等(2018)	283/162:121/6~24 mon	0.62	0.81	0.67	0.78
小计: 10 项工具	21 项研究	4047*/3101:1937(4 项研究性别有缺失值)/15.62–26.94 mon ($M = 20.62$)	0.72	0.88	0.59	0.94

注: * 4047 是表 3 中婴儿阶段 10 项筛查工具所涉及样本数量的平均值。其中 23966 是使用 M-CHAT 这一筛查工具的 4 项研究样本量的平均值, 包括 Carbone 等(2020)的样本数量是 26364, Koh 等(2014)的样本数量是 17302, Stenberg 等(2014)的样本数量是 52026, Toh 等(2018)的样本数量是 173。5272 是使用 M-CHAT/F 这一筛查工具的 6 项研究样本量的平均值, 包括 Baduel 等(2017)的样本数量是 1250, Guthrie 等(2019)的样本数量是 25999, Kamio 等(2014)的样本数量是 1851, Kerub 等(2020)的样本数量是 1591, Srisinghasongkram 等(2016)的样本数量是 841, Sturner 等(2016)的样本数量是 98。6886 是使用 M-CHAT-R/F 这一筛查工具的 4 项研究样本量的平均值, 包括 Coelho-Medeiros 等(2019)的样本数量是 120, Guo 等(2019)的样本数量是 7928, Magán-Maganto 等(2020)的样本数量是 3426, Robins 等(2014)的样本数量是 16071。

表 4 幼儿阶段自闭症风险筛查工具的分类准确性

工具名称	研究者，年代	样本数量/性别(男:女)/年龄(平均值)		敏感性	特异性	阳性预测值 PPV	阴性预测值 NPV
M-CHAT	Koh 等(2014) Toh 等(2018)	1201/655:546/28.5~42 mon (M = 35.25)		0.70	0.86	0.69	0.92
M-CHATR/F	Magán-Maganto 等(2020) Tsai 等(2019)	1703/878:825/19.5~34 mon (M = 24.39)		0.81	0.97	0.42	0.99
CBCL1½–5	Havdahl 等(2016) Limberg 等(2017) Narzisi 等(2013)	162/119:43/22.33~55.67 mon (M = 44.97)	退缩	0.85	0.71	0.82	0.90
			PDP	0.91	0.76	0.77	0.90
RC_ASRS	Zhou 等(2018)	1918/1056:862/24~60 mon (M = 43.8)		0.886	0.845	NR	NR
Q-CHAT	Ruta 等(2019)	315/206:109/M = 31.6 mon		0.8273	0.7778	NR	NR
TIDOS	Oner 等(2014)	259/179:80/18~60 mon (M = 38)		0.95	0.91	0.85	0.98
T-STAT	Chiang 等(2013)	30/20:10/24~35 mon (M = 28.41)		0.93	0.87	0.88	0.93
ADEC	Hedley 等(2015)	96/76:20/12~36 mon (M = 28.67)		0.93	0.64	0.78	0.88
AMSE	Cederlund (2019)	126/96:30/M = 54.4 mon		0.75	0.78	0.95	0.35
OERA	Paula 等(2018)	99/76:23/36~120 mon (M = 62.04)		0.9275	0.9091	NR	NR
RITA-T	Choueiri 和 Wagner (2015)	61/41:20/18~36 mon (M = 26.1)		1	0.84	0.88	1
BISCUIT-Part1	Horovitz 和 Matson (2014)	2048/1461:576(11 人性别不明)/27.50~33.50 mon (M = 29.87)		0.87	0.87	NR	NR
DCI	Janvier 等(2019)	362/249:113/24~60 mon (M = 38)		0.66	0.76	NR	NR
PAAS	Perera 等(2017)	105/77:28/18~48 mon (M = 36.16)		0.880	0.933	0.952	0.840
小计：14 项工具	18 项研究(4 项同婴儿阶段)	606*/371:235/22.36~51.97 mon (M = 37.26)		0.86	0.83	0.80	0.87

注：* 606 是表 4 中幼儿阶段 14 项筛查工具所涉及样本数量的平均值。其中 1201 是使用 M-CHAT 这一筛查工具的 2 项研究样本量的平均值，包括 Koh 等(2014)的样本数量是 407, Toh 等(2018)的样本数量是 1995。1703 是使用 M-CHAT/F 这一筛查工具的 2 项研究样本量的平均值，包括 Magán-Maganto 等(2020)的样本数量是 3089, Tsai 等(2019)的样本数量是 317。162 是使用 CBCL1½–5 这一筛查工具的 3 项研究样本量的平均值，包括 Havdahl 等(2016)的样本数量是 161，Limberg 等(2017)的样本数量是 183, Narzisi 等(2013)的样本数量是 141。

而婴儿阶段特别是低于 12 个月的筛查工具却很少？可以从两个方面来分析这一问题，一方面，36 个月(3 岁)之前发生是自闭症诊断的关键年龄，随着年龄的增长，自闭症的症状特征也日趋明显，因此，对于 3~6 岁幼儿的筛查工具的开发和使用已经为全世界范围所共识，这一阶段筛查工具的数量和质量也是成正比例的。另一方面，对于 24 个月(2 岁)之前，特别是 12 个月(1 岁)的自闭症风险预警的行为指标尚不明确。尽管有文献表明，许多儿童在 12 个月前可能出现 ASD 的危险迹象(Zwaigenbaum et al., 2005)，但其他研究报告也称，35%至 46%的自闭症儿童有“迟发”症状，这些发现指出，很早就进行 ASD 风险筛查，但后来并没有发展成 ASD 儿童(Landa et al., 2007; Werner et al., 2005)，也有一些前瞻性研究表明，自闭症儿童在 12 个月后可能会出现很大比例的发育退化(Ozonoff et al., 2010)。因此，12 个月大时漏诊病例的可能解释是(a)一些儿童没有出现 ASD 行为指标，(b)研究人员尚未确定 12 个月大时预测 ASD 最终诊断的所有关键行为指标，(c)研究人员尚未找到基于家长报告的能测量儿童关键行为的可靠指标。但是，需要强调的是，对于低于 12 个月筛查为阳性或阴性的儿童都不应仅视为单点事件(single-point screening event)，需要开展更多的基于婴儿家庭录像(home video)的回顾性分析和高危儿童的前瞻性研究，重点探讨那些当单点筛查被认为有/无风险的儿童，至最后成为自闭症谱系/常态儿童，在其发展轨迹过程中出现了哪些问题，以确定自闭症在社交、交流、行为、运动和气质等领域可能是危险迹象的其他行为，并帮助父母

在 12 个月之前识别和确定自己婴儿的这些行为。

其次, 在筛查工具的心理测量学方面, 综述中所涉及到 35 项研究中只有几个心理测量特性比其他特性更常使用, 如评估内部一致性信度、评分者信度、重测信度和效标效度; 而分半信度、测量误差、内容效度、结构效度、跨文化效度和假设检验只在少数研究中出现, 有 1/3 的研究根本没有开展任何心理测量学评估, 这也是造成心理测量特性存在高风险偏倚的主要原因之一。还有一个非常常见的问题是缺失数据的处理。很少有作者明确地量化了自己研究数据集中缺失的数据, 也没人解释如何处理缺失数据, 都遵循了哪些方法? 例如, 对于基于父母报告的测量, 有可能有一些项目是没有答案的, 对于这种特殊情况, 一般通过统计程序插补数据, 但这种处理有可能改变数据结构和分布, 进而导致对自闭症风险的高估或低估。因此, 对于主旨在于识别自闭症风险早期迹象的筛查研究来说, 缺失数据的处理是一个至关重要的方面, 在未来研究中, 研究人员需要解释他们是否以及如何处理样本中缺失的数据。总的说来, 从我们对筛查工具的心理测量特性的检验结果可以得出两个主要的考虑因素, 一个与心理测量评估的数量有关, 另一个与研究本身方法学质量有关。而且要着重关注的是第二个因素, 与其他研究相比, 对筛查工具的研究普遍缺乏内容效度的评估, 少量研究采用假设检验来评估工具的区分度, 但得到方差分析的结果较差或缺少跨研究进行衡量的统一标准。这些发现将推动研究人员设计验证研究, 重点要放在提升筛查工具的测量学质量上。

第三, 在文化适应性上, 尽管金标准的诊断工具 ADOS 和 ADI-R 已被翻译成多种语言, 并在不同的跨国环境中进行了验证, 但使用这些工具需要花费大量时间和金钱。因此, 在低收入和中等收入国家中, 金标准诊断工具的使用明显滞后, 而常采用具有广泛文化适应性验证的筛查工具, 如 M-CHAT 系列工具(包括 M-CHAT, M-CHAT/F, M-CHAT-R/F), 经过国际上广泛应用和多种语言中进行评估, M-CHAT 系列工具已经表现出明显的文化适应性优势(Guo et al., 2019; Inada et al., 2011; Seif Eldin et al., 2008), 被誉为最广泛使用和最著名的父母报告筛查工具。但值得的注意的是, 在低资源的社区或国家背景下, 开发和使用适合其当地文化和语言的筛查工具, 这一趋势也正在增加(Stewart & Lee, 2017)。本研究关于自闭症风险筛查工具的文化适应性结果显示, 大部分验证性研究对于文化适应性的报道较为缺乏, 严格控制的程度也有很大差异。因此, 我们建议未来研究需要更全面地检查文化适应性程序(至少要包括翻译、回译、委员会审查和实地预测试), 同时还需要解决的重要问题是, 涉及文化或语言上不同的亚群体, 或识字率较低或没有识字能力的人, 有必要使用经过培训的双文化/双语个体来管理筛查工具, 以产生充分的理解。

在筛查工具的编制和诊断结果符合诊断标准方面, 国际最好的自闭症谱系诊断方法是基于国际疾病分类(ICD)和美国精神障碍诊断和统计手册(DSM)的框架, 使用专家临床判断作为确认诊断的“金标准”。尽管我们检索的是从 2013 年 DSM-5 发布到 2020 年之间的文献, 部分自闭症风险筛查测量中使用的诊断标准已经最大限度地遵循了 DSM-5, 涵盖了 DSM-5 手册中的 A 类(社会互动和社会沟通)和 B 类(限制性和重复性行为)中的至少两个标准。但是, DSM-IV 或 DSM-IV-TR 的参考和使用仍然占比最多, 包括自闭症诊断观察表(ADOS), 以及半结构化的家长访谈, 如自闭症诊断访谈−修订版(ADI-R)也是在 2013 年之前制定的, 也没有考虑 DSM-5 中制定的指南。由于一些使用 DSM-IV 标准诊断为自闭症的患者可能不属于 DSM-5 的自闭症患者, 所以在行为科学、精神病学和心理学领域中, 关于 DSM-IV 标准(包括修订版)和 DSM-5 标准优越性的争论也正日趋白热化, 如 Grzadzinski 等人(2013)研究显示 DSM-5 对成人和幼儿的敏感性降低, 而 Sappok 等人(2015)则显示在 DSM-IV 和 DSM-5 指导下的筛查测量病例的敏感性是相一致的, 尽管特异性有所下降。因此, 有必要重新检查 ASD 诊断和筛查工具中的问题和特征, 以全面满足 DSM-5 的新标准。这就需要将新的 ASD 标准映射到筛查工具中使用的项目中, 同时还要评估诊断过程的工作方式, 这种尝试有可能导致当前筛查工具版本的更新。此外, 还需要进行综合实验研究, 使用对照组和病例作为数据, 以便指导研究人员、临床医生、精神病医生和心理学家使用正确的筛选工具, 即使在提出 DSM-5 新的变化之后也能保持绩效。

尽管从研究和临床的角度来看, 自闭症早期

发现的重要性也已经达成共识，但是选择一种适合特定背景的筛查程序仍然困难重重，无论是对婴幼儿进行指定时间点的特异性筛查，或者面向健康婴幼儿没有顾虑的广普性筛查，无论是通过发育监测和使用正式筛查相结合的方式，或是在特定年龄使用单一筛查并在不同年龄段重复监测，任何选择除了参考如上应用水平、心理测量特性、文化适应性、符合诊断标准等有效性指标以外，至关重要的一个问题是对筛查结果应该谨慎解释，特别是临床医生要了解如何解释已经发表的研究数据，如要考虑到由于阴性病例没有得到随访可能会扭曲筛查结果，识别漏诊病例的可能性等都应仔细告知家长，以确保家庭纵向参与和后续治疗。我们建议采用一种更广泛、更平衡的方法来说明筛查结果，可以由具有 ASD 专门知识的多学科儿童专家小组(一般包括一名能够提供医疗诊断的医学专家或临床心理学家，及言语病理学家、职业治疗师、医务社会工作者等)完成，更可能考虑到影响筛查结果的所有因素，这一多学科团队共识诊断的诊断准确性(即随时间的稳定)估计为 80–90% (Woolfenden et al., 2012)。总之，规范筛查工具的质量评估标准，在政策支持的背景下，开展这些筛查管理和实用工具的临床试验，以优化我们识别自闭症谱系儿童的能力，并制定一致性的策略，以便在以后的年龄持续地监测自闭症症状。

4.2 对照分类准确性分析数据，批判性选择自闭症谱系风险筛查工具

第二个研究问题是我们希望根据 QUADAS-2 的评估结果，确定一个(或多个)用于评估自闭症的早期风险迹象的有效筛查工具。我们重点检视 35 篇文献信息来源，其中有 27 项研究采用由父母或主要照顾者完成的问卷或清单来收集信息和数据，占比 77%，如 M-CHAT 系列检查表。这主要是由问卷本身所固有的优点所致。首先，问卷调查所需时间很短，不需要特定的知识或培训，侵入性要比观察检查表或访谈小得多。其次，通常不需要对问卷的编码系统或分数解释进行专门培训。对于许多问卷来说，最终得分的插补和意义的归纳也不涉及任何 ASD 的临床解释或具体知识。然而，调查问卷有几个局限性。首先，分数取决于信息提供者的主观性。由于调查问卷主要是为父母设计的，但父母的回顾性报告是缺乏准确性的，他们可能会低估或高估风险的早期迹象，这取决于他们发现风险的能力，以及区分风险迹象与正常偏离发展轨迹的能力。问卷的另一个固有局限性是社会期望偏差，表现为过度报告期望行为。

对比来看，只有 8 项研究(占比 23%)采用观察法，以二级筛查测量为多见。虽然观察法的优点特别明显，如能够获得更深入，真实的资料；适合对一些不能使用问卷、访谈法的研究对象，如婴儿的个体行为、活动的研究，但是受限于对低幼儿童的行为观察是一种劳动密集型的任务，观察者只能捕获有限的行为样本，而且还需要考虑到儿童在观察时的反应性，考虑到儿童与父母或临床医生之间的互动具有动态性，易出现霍桑效应等，这些来自观察员、观察方式、观察时间和地点的影响都会对观察质量产生干扰。因此，对观察者进行统一的专业培训是减少研究误差，确保观察信度的重要手段，而这也正是观察法的主要局限。

统观问卷法和观察法的局限性，我们发现“改良的幼儿自闭症检查表−有修订的后续随访”(M-CHAT-R/F)这一筛查工具似乎可以有效解决这一困境，当家长填写修订后的 20 项 M-CHAT 问卷，筛查结果呈阳性，需要进一步参与结构化的随访访谈和/或面谈，在一项针对 16 至 31 个月大的幼儿的大样本美国研究中，使用 M-CHAT-R/F 比 M-CHAT/F 识别出更多的自闭症儿童(67 例/10000 比 45 例/10000)，$\chi^2 = 8.63$; $p = 0.003$ (Chlebowski et al., 2013)。

结合分类准确性分析的数据结果，我们发现，M-CHAT-R/F 分别在婴儿阶段和幼儿阶段也保持了良好的辨别水平，这也表明其可能是目前最具潜力的自闭症谱系筛查工具之一。值得注意的是，虽然如上这些筛查工具都很有希望，但目前并没有一项可以被认为是早期发现 ASD 风险的金标准，这一领域的研究和发展还有很大的进取空间。

5 研究局限与展望

本综述有两个局限性。首先，我们采用的 QUADAS-2 是一种评估研究方法学质量的标准化方案，而不是评估筛查工具本身。因此对 QUADAS-2 评估方案自身质量的信心是确定其评估结果是否可靠和可信的第一步，换言之，评估

方案中所划定的评价标准的数量和质量可以决定评价结果中存在偏差的水平高低。例如, 我们的质量评价标准中没有考虑“研究类型” (根据筛查研究的研究性质, 可分为“验证研究、标准化测量、跨文化比较、纵向或后续研究”4 个子指标)这一审查维度, 这也导致后期文献梳理过程中无法根据“文化适应性”的指标来全面衡量每一项研究。同时, 本综述还缺乏对于“流程和时间”方面的偏倚风险评价指标。因此, 对于应用 QUADAS-2 来评估研究的测量特性, 进而评估筛查工具的测量特性, 我们建议未来的筛查工具研究需要采取严格的质量分析审查方案, 如包括一个流程图, 介绍关于被试招募方法、样本、测试执行顺序、随访以及与过程相关的其他细节的信息, 以提高可复制性, 并更好地告知读者潜在的偏倚。

第二, QUADAS-2 的方案采用的是最差计分原则, 且缺少可降低论文异质性数据分析的统计学分析, 这些都可能会导致对某些工具使用的负面偏见。由于 QUADAS-2 本身的特点在于其方案中的每一项都是整体评估的重要组成部分, 因此任何项目的差评都应被视为严重缺陷。例如, 我们将样本量⩾100 为优, 50~99 为良, 30~49 为可, < 30 为差(Terwee et al., 2012)。对于一般人群时, 这种分类是一个很好的标准, 而当考虑到自闭症风险和/或临床组时, 则应根据 ASD 的患病率仔细考虑这一样本量划分指标可能并不适当。另外, 对分类准确性的评价指标, 我们只考虑了敏感性和特异性的合并平均值进行每一项工具的横向比较, 而根据层次综合被试工作特征曲线(Hierarchical Summary Receive Operating Charateristic, HSROC)模型, 敏感性和特异性之间可能是非线性的关系(Sánchez-García et al., 2019), 可以采用贝叶斯层次模型来调整自闭症筛查工具的不完善性(Rutter & Gatsonis, 2001), 可以更严格地检查抽样的变异性和纳入论文之间的异质性, 以确保分析结果的准确性, 这些问题应被视为当前研究的局限性。

在过去的七年里, 与自闭症症状描述及筛查相关的研究呈上升趋势。本综述以上列出的是常见的 ASD 筛查工具, 未来研究仍然需要许多利益相关者的参与和有效合作, 增强对筛查过程进行深度的比较, 如继续搜索有辨别力的筛查项目内容或行为、参与筛查预测试、计算分数、解释和解释测试结果的标准化, 如何处理进一步的转介, 时间消耗和 NPV 性能等, 以进一步确定性能最好的工具。其次, 在筛查工具的引进和管理上, 面对国内 ASD 早期筛查与诊断技术仍比较滞后的现状, 在鼓励引进国际通用的自闭症诊断工具 ADOS 和 ADI-R 的同时, 支持自主研发符合国际标准(ICD-11、DSM-5)的本土化自闭症诊断工具, 更要加大对公共卫生学、医学心理学、儿科保健等多学科专业人员的培训力度, 完善婴幼儿的转介系统职能, 扩大专门小组进行早期诊断和康复干预的能力。第三个建议是一个受到 Covid- 19 的传播而变得突出的现实问题, 病毒的传播及其毁灭性的后果导致了公共卫生社区服务处于隔离状态、父母怀疑孩子有发育问题时寻求专业意见的可能性下降、面对面康复服务可及性降低, 因此, 开发和使用基于机器学习智能算法的新型 ASD 的自我管理工具, 如自闭症指纹(Autism Fingerprint) (Klein et al., 2015)、自闭症谱系障碍筛查 APP (ASD Tests) (Thabtah, 2019), 通过使用机器学习算法生成的分类器计算筛查的敏感性和特异性来保障筛查的准确性, 由专业医疗中心发布的医疗报告来确诊正规诊断状态来实现远程保健服务, 这些基于移动应用程序的筛查方式不仅为家庭提供一个可靠和易于使用的工具来筛查他们的孩子是否患有自闭症, 更有助于用户和健康社区克服隔离状态的困境。

参考文献

*表示元分析用到的文献

陈文雄. (2013). 孤独症 70 年: 从 kanner 到 dsm-v. *临床儿科杂志, 31*(11), 1001–1004.

方俊明. (2005). *特殊教育学*. 北京：人民教育出版社.

冯雅静, 王雁. (2012). 孤独症儿童的诊断工具:现状及展望. *中国特殊教育*, (9), 45–52+70.

龚郁杏, 刘靖, 郭延庆, 宋文红, 贾美香, 李长璟. (2015). 改良婴幼儿孤独症量表中文简化版的效度和信度. *中国心理卫生杂志, 29*(2), 121–124.

关智勇, 龚建华, 陈艳琳, 周首俊, 李丽. (2016). 儿童孤独症谱系测验量表的信度、效度及临界值. *中华实用儿科临床杂志, 31*(11), 855–858.

何华国. (2006). *特殊幼儿早期疗育*. 台中：五南图书出版股份有限公司.

骆名进, 宋海东, 刘健. (2020). 基于社区开展家长自评式孤独症早期筛查研究. *中国全科医学*, *23*(17), 2219–2222.

朴永馨. (2014). *特殊教育辞典* (pp. 3–4). 华夏出版社.

徐建平, 张厚粲. (2005). 质性研究中编码者信度的多种方法考察. *心理科学, 28*(6), 1430–1432.

王辉. (2015). *特殊儿童教育评估*. 南京：南京师范大学出版社.

王辉, 李晓庆, 李晓娟. (2009). 国内孤独症儿童评估工具的研究现状. *中国特殊教育,* (7), 54–59+43.

魏华忠. (1995). *教育评价与技术*. 大连：辽宁师范大学出版社.

许丹, 刘黎虹, 林妙莲, 宋海东. (2013). 儿童孤独症测验的应用研究. *中国儿童保健杂志, 21*(12), 1259–1261.

杨清. (1985). *简明心理学辞典* (p. 94, 316). 吉林人民出版社.

张福娟, 贺莉. (2001). 自闭症儿童的诊断与评估. *现代康复,* (11), 100–101.

张永盛, 吕超. (2013). 自闭症谱系障碍儿童诊断评估概况——自闭症儿童认知加工及干预研究. *绥化学院学报,* (4), 116–121.

周念丽, 方俊明. (2008). 自闭症谱系障碍儿童心理测评的回溯与探索. *心理科学,* (6), 1330–1333+1312.

*Baduel, S., Guillon, Q., Afzali, M. H., Foudon, N., Kruck, J., & Rogé, B. (2017). The French version of the modified-checklist for autism in toddlers (M-CHAT): A validation study on a French sample of 24 month-old children. *Journal of Autism and Developmental Disorders*, *47*(2), 297–304.

*Carbone, P. S., Campbell, K., Wilkes, J., Stoddard, G. J., Huynh, K., Young, P. C., & Gabrielsen, T. P. (2020). Primary care autism screening and later autism diagnosis. *Pediatrics*, *146*(2), Article e20192314. https://doi.org/10.1542/peds.2019-2314

CDC. (2012). Prevalence of autism spectrum disorders—autism and developmental disabilities monitoring network, 14 sites, United States, 2008. *Morbidity and Mortality Weekly Report: Surveillance Summaries, 61*(3), 1–19.

CDC. (2014). Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *Morbidity and Mortality Weekly Report: Surveillance Summaries, 63*(2), 1–21.

*Cederlund, M. (2019). Autism Mental Status Examination (AMSE): A valid instrument in the evaluation of pre-school children with suspected autism spectrum disorders? *Journal of Autism and Developmental Disorders*, *49*(7), 2965–2979.

Charman, T., & Gotham, K. (2013). Measurement issues: Screening and diagnostic instruments for autism spectrum disorders–lessons from research and practise. *Child and Adolescent Mental Health*, *18*(1), 52–63.

*Chiang, C. H., Wu, C. C., Hou, Y. M., Chu, C. L., Liu, J. H., & Soong, W. T. (2013). Development of T-STAT for early autism screening. *Journal of Autism and Developmental Disorders*, *43*(5), 1028–1037.

Chlebowski, C., Robins, D. L., Barton, M. L., & Fein, D. (2013). Large-scale use of the modified checklist for autism in low-risk toddlers. *Pediatrics*, *131*(4), e1121–e1127.

*Choueiri, R., & Wagner, S. (2015). A new interactive screening test for autism spectrum disorders in toddlers. *The Journal of Pediatrics*, *167*(2), 460–466.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290.

Cicchetti, D. V., Volkmar, F., Klin, A., & Showalter, D. (1995). Diagnosing autism using ICD-10 criteria: A comparison of neural networks and standard multivariate procedures. *Child Neuropsychology*, *1*(1), 26–37.

*Coelho-Medeiros, M. E., Bronstein, J., Aedo, K., Pereira, J. A., Arraño, V., Perez, C. A., ... Bedregal, P. (2019). M-CHAT-R/F validation as a screening tool for early detection in children with autism spectrum disorder. *Revista Chilena de Pediatria*, *90*(5), 492–499.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., ... Varley, J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model. *Pediatrics*, *125*(1), 17–23.

Falkmer, T., Anderson, K., Falkmer, M., & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: A systematic literature review. *European Child & Adolescent Psychiatry*, *22*(6), 329–340.

García-Primo, P., Hellendoorn, A., Charman, T., Roeyers, H., Dereu, M., Roge, B., ... Canal-Bedia, R. (2014). Screening for autism spectrum disorders: State of the art in Europe. *European Child & Adolescent Psychiatry, 23*(11), 1005–1021.

Grønborg, T. K., Schendel, D. E., & Parner, E. T. (2013). Recurrence of autism spectrum disorders in full- and half-siblings and trends over time: A population-based cohort study. *JAMA Pediatrics, 167*(10), 947–953.

Grzadzinski, R., Huerta, M., & Lord, C. (2013). DSM-5 and autism spectrum disorders (ASDs): An opportunity for identifying ASD subtypes. *Molecular Autism*, *4*(1), 1–6.

Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology*, *46*(12), 1417–1432.

*Guo, C., Luo, M., Wang, X., Huang, S., Meng, Z., Shao, J., ... Jing, J. (2019). Reliability and validity of the Chinese version of modified checklist for autism in toddlers, revised, with follow-up (M-CHAT-R/F). *Journal of Autism and Developmental Disorders*, *49*(1), 185–196.

*Guthrie, W., Wallis, K., Bennett, A., Brooks, E., Dudley, J., Gerdes, M., ... Miller, J. S. (2019). Accuracy of autism screening in a large pediatric network. *Pediatrics*, *144*(4), Article e20183963. https://doi.org/10.1542/peds.2018-3963

Hampton, J., & Strand, P. S. (2015). A review of level 2 parent-report instruments used to screen children aged 1.5–5 for autism: A meta-analytic update. *Journal of Autism and Developmental Disorders*, *45*(8), 2519–2530.

*Havdahl, K. A., von Tetzchner, S., Huerta, M., Lord, C., & Bishop, S. L. (2016). Utility of the child behavior checklist as a screener for autism spectrum disorder. *Autism Research*, *9*(1), 33–42.

*Hedley, D., Nevill, R. E., Monroy-Moreno, Y., Fields, N., Wilkins, J., Butter, E., & Mulick, J. A. (2015). Efficacy of the ADEC in identifying autism spectrum disorder in clinically referred toddlers in the US. *Journal of Autism and Developmental Disorders*, *45*(8), 2337–2348.

*Horovitz, M., & Matson, J. L. (2014). The baby and infant screen for children with autism traits-part 1: Age-based scoring procedures. *Journal of Developmental and Physical Disabilities*, *26*(1), 1–22.

Inada, N., Koyama, T., Inokuchi, E., Kuroda, M., & Kamio, Y. (2011). Reliability and validity of the Japanese version of the modified checklist for autism in toddlers (M-CHAT). *Research in Autism Spectrum Disorders*, *5*(1), 330–336.

*Janvier, Y. M., Coffield, C. N., Harris, J. F., Mandell, D. S., & Cidav, Z. (2019). The Developmental Check-In: Development and initial testing of an autism screening tool targeting young children from underserved communities. *Autism*, *23*(3), 689–698.

*Kamio, Y., Inada, N., Koyama, T., Inokuchi, E., Tsuchiya, K., & Kuroda, M. (2014). Effectiveness of using the modified checklist for autism in toddlers in two-stage screening of autism spectrum disorder at the 18-month health check-up in Japan. *Journal of Autism and Developmental Disorders*, *44*(1), 194–203.

*Kerub, O., Haas, E. J., Meiri, G., Davidovitch, N., & Menashe, I. (2020). A comparison between two screening approaches for ASD among toddlers in Israel. *Journal of Autism and Developmental Disorders*, *50*(5), 1553–1560.

Klein, T. J., Al-Ghasani, T., Al-Ghasani, M, Akbar, A., Tang, E., & Al-Farsi, Y. (2015). A mobile application to screen for autism in Arabic-speaking communities in Oman. *The Lancet Global Health, 3,* S15.

*Koh, H. C., Lim, S. H., Chan, G. J., Lin, M. B., Lim, H. H., Choo, S. H. T., & Magiati, I. (2014). The clinical utility of the modified checklist for autism in toddlers with high risk 18–48 month old children in Singapore. *Journal of Autism and Developmental Disorders*, *44*(2), 405–416.

Landa, R. J., Holman, K. C., & Garrett-Mayer, E. (2007). Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders. *Archives of General Psychiatry*, *64*(7), 853–864.

Lecciso, F., Petrocchi, S., Savazzi, F., Marchetti, A., Nobile, M., & Molteni, M. (2013). The association between maternal resolution of the diagnosis of autism, maternal mental representations of the relationship with the child, and children's attachment. *Life Span Disabil, 16*(1), 21–38.

Le Couteur, A., Haden, G., Hammal, D., & McConachie, H. (2008). Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments: The ADI-R and the ADOS. *Journal of Autism and Developmental Disorders*, *38*(2), 362–372.

Leo, M., Carcagnì, P., Distante, C., Mazzeo, P. L., Spagnolo, P., Levante, A., ... Lecciso, F. (2019). Computational analysis of deep visual data for quantifying facial expression production. *Applied Sciences, 9*(21), 4542.

Levy, S. E., Wolfe, A., Coury, D., Duby, J., Farmer, J., Schor, E., ... Warren, Z. (2020). Screening tools for autism spectrum disorder in primary care: A systematic evidence review. *Pediatrics*, *145*(Suppl. 1), S47–S59.

*Limberg, K., Gruber, K., & Noterdaeme, M. (2017). The German version of the child behavior checklist 1.5–5 to identify children with a risk of autism spectrum disorder. *Autism*, *21*(3), 368–374.

*Magán-Maganto, M., Canal-Bedia, R., Hernández-Fabián, A., Bejarano-Martín, Á., Fernández-Álvarez, C. J., Martínez-Velarte, M., ... de la Paz, M. P. (2020). Spanish cultural validation of the modified checklist for autism in toddlers, revised. *Journal of Autism and Developmental Disorders*, *50*(7), 2412–2423.

Mandell, D. S., Morales, K. H., Xie, M., Lawer, L. J., Stahmer, A. C., & Marcus, S. C. (2010). Age of diagnosis among Medicaid-enrolled children with autism, 2001–2004. *Psychiatric Services, 61*(8), 822–829.

Marlow, M., Servili, C., & Tomlinson, M. (2019). A review of screening tools for the identification of autism spectrum disorders and developmental delay in infants and young children: Recommendations for use in low‐and middle-income countries. *Autism Research*, *12*(2), 176–199.

McDonald, S. D., Brown, W. L., Benesek, J. P., & Calhoun, P. S. (2015). A systematic review of the PTSD checklist's diagnostic accuracy studies using QUADAS. *Psychological Trauma: Theory, Research, Practice, and Policy*, *7*(5), 413–421.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews

and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 6*(7), e1000097.

*Narzisi, A., Calderoni, S., Maestro, S., Calugi, S., Mottes, E., & Muratori, F. (2013). Child behavior check list 1½–5 as a tool to identify toddlers with autism spectrum disorders: A case-control study. *Research in Developmental Disabilities*, *34*(4), 1179–1189.

Norris, M., & Lecavalier, L. (2010). Screening accuracy of level 2 autism spectrum disorder rating scales: A review of selected instruments. *Autism, 14*(4), 263–284.

*Oner, P., Oner, O., & Munir, K. (2014). Three-item direct observation screen (TIDOS) for autism spectrum disorder. *Autism*, *18*(6), 733–742.

Ozonoff, S., Iosif, A. M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., ... Young, G. S. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(3), 256–266.

*Paula, C. S., Cunha, G. R., Bordini, D., Brunoni, D., Moya, A. C., Bosa, C. A., ... Cogo-Moreira, H. (2018). Identifying autism with a brief and low-cost screening instrument—OERA: Construct validity, invariance testing, and agreement between judges. *Journal of Autism and Developmental Disorders*, *48*(5), 1780–1791.

*Perera, H., Jeewandara, K. C., Seneviratne, S., & Guruge, C. (2017). Culturally adapted pictorial screening tool for autism spectrum disorder: A new approach. *World Journal of Clinical Pediatrics*, *6*(1), 45–51.

*Raza, S., Zwaigenbaum, L., Sacrey, L. A. R., Bryson, S., Brian, J., Smith, I. M., ... Garon, N. (2019). Brief report: Evaluation of the short quantitative checklist for autism in toddlers (Q-CHAT-10) as a brief screen for autism spectrum disorder in a high-risk sibling cohort. *Journal of Autism and Developmental Disorders*, *49*(5), 2210–2218.

Reichow, B., Barton, E. E., Boyd, B. A., & Hume, K. (2012). Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD). *Cochrane Database of Systematic Reviews, 5*(5), CD009260. https://doi.org/10.1002/14651858.CD009260.pub2

Renty, J., & Roeyers, H. (2006). Satisfaction with formal support and education for children with autism spectrum disorder: The voices of the parents. *Child: Care, Health and Development, 32*(3), 371–385.

*Rescorla, L. A., Winder-Patel, B. M., Paterson, S. J., Pandey, J., Wolff, J. J., Schultz, R. T., & Piven, J. (2019). Autism spectrum disorder screening with the CBCL/1½–5: Findings for young children at high risk for autism spectrum disorder. *Autism*, *23*(1), 29–38.

*Robins, D. L., Casagrande, K., Barton, M., Chen, C. M. A., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics*, *133*(1), 37–45.

Robins, D. L., & Dumont-Mathieu, T. M. (2006). Early screening for autism spectrum disorders: Update on the modified checklist for autism in toddlers and other measures. *Journal of Developmental & Behavioral Pediatrics*, *27*(2), S111–S119.

*Ruta, L., Chiarotti, F., Arduino, G. M., Apicella, F., Leonardi, E., Maggio, R., ... Muratori, F. (2019). Validation of the quantitative checklist for autism in toddlers in an Italian clinical sample of young children with autism and other developmental disorders. *Frontiers in Psychiatry*, *10*, 488.

Rutter, C. M., & Gatsonis, C. A. (2001). A hierarchical regression approach to meta‐analysis of diagnostic test accuracy evaluations. *Statistics in Medicine*, *20*(19), 2865–2884.

*Sacrey, L. A. R., Bryson, S., Zwaigenbaum, L., Brian, J., Smith, I. M., Roberts, W., ... Garon, N. (2018). The autism parent screen for infants: Predicting risk of autism spectrum disorder based on parent-reported behavior observed at 6–24 months of age. *Autism*, *22*(3), 322–334.

Sánchez-García, A. B., Galindo-Villardón, P., Nieto-Librero, A. B., Martín-Rodero, H., & Robins, D. L. (2019). Toddler screening for autism spectrum disorder: A meta-analysis of diagnostic accuracy. *Journal of Autism and Developmental Disorders*, *49*(5), 1837–1852.

Sappok, T., Heinrich, M., & Underwood, L. (2015). Screening tools for autism spectrum disorders. *Advances in Autism, 1*(1), 12–29.

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). San Diego: Author.

Scott, F. J., Baron-Cohen, S., Bolton, P., & Brayne, C. (2002). The CAST (Childhood Asperger Syndrome Test): Preliminary Development of a UK Screen for Mainstream Primary-School-Age Children. *Autism, 6*(1), 9–31.

Seif Eldin, A., Habib, D., Noufal, A., Farrag, S., Bazaid, K., Al-Sharbati, M., ... Gaddour, N. (2008). Use of M-CHAT for a multinational screening of young children with autism in the Arab countries. *International Review of Psychiatry*, *20*(3), 281–289.

Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology, 70*(1), 747–770.

Soleimani, F., Khakshour, A., Abassi, Z., Khayat, S., Ghaemi, S. Z., Azam, N., & Hajikhani Golchin, N. A. (2014). Review of autism screening tests. *International Journal of Pediatrics*, *2*(4.1), 319–329.

Soto, S., Linas, K., Jacobstein, D., Biel, M., Migdal, T., & Anthony, B. J. (2015). A review of cultural adaptations of screening tools for autism spectrum disorders. *Autism*, *19*(6), 646–661.

*Srisinghasongkram, P., Pruksananonda, C., & Chonchaiya, W. (2016). Two-step screening of the modified checklist for autism in toddlers in Thai children with language delay and typically developing children. *Journal of Autism and Developmental Disorders*, *46*(10), 3317–3329.

*Stenberg, N., Bresnahan, M., Gunnes, N., Hirtz, D., Hornig, M., Lie, K. K., ... Stoltenberg, C. (2014). Identifying children with autism spectrum disorder at 18 months in a general population sample. *Paediatric and Perinatal Epidemiology, 28*(3), 255–262.

Stewart, L. A., & Lee, L. C. (2017). Screening for autism spectrum disorder in low-and middle-income countries: A systematic review. *Autism*, *21*(5), 527–539.

Stout, M. J., Conner, S. N., Colditz, G. A., Macones, G. A., & Tuuli, M. G. (2015). The utility of 12-hour urine collection for the diagnosis of preeclampsia: A systematic review and meta-analysis. *Obstetrics & Gynecology*, *126*(4), 731–736.

*Sturner, R., Howard, B., Bergmann, P., Morrel, T., Andon, L., Marks, D., ... Landa, R. (2016). Autism screening with online decision support by primary care pediatricians aided by M-CHAT/F. *Pediatrics*, *138*(3), Article e20153036. https://doi.org/10.1542/peds.2015-3036

Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*, *21*(4), 651–657.

Thabtah, F. (2019). An accessible and efficient autism screening method for behavioural data and predictive analyses. *Health Informatics Journal, 25*(4), 1739–1755

Thabtah, F., & Peebles, D. (2019). Early autism screening: A comprehensive review. *International Journal of Environmental Research and Public Health*, *16*(18), 3502.

*Toh, T. H., Tan, V. W. Y., Lau, P. S. T., & Kiyu, A. (2018). Accuracy of modified checklist for autism in toddlers (M-CHAT) in detecting autism and other developmental disorders in community clinics. *Journal of Autism and Developmental Disorders*, *48*(1), 28–35.

*Tsai, J. M., Lu, L., Jeng, S. F., Cheong, P. L., Gau, S. S. F., Huang, Y. H., & Wu, Y. T. (2019). Validation of the modified checklist for autism in toddlers, revised with follow-up in Taiwanese toddlers. *Research in Developmental Disabilities*, *85*, 205–216.

*Turner-Brown, L. M., Baranek, G. T., Reznick, J. S., Watson, L. R., & Crais, E. R. (2013). The first year inventory: A longitudinal follow-up of 12-month-old to 3-year-old children. *Autism*, *17*(5), 527–540.

Volkmar, F., Siegel, M., Woodbury-Smith, M., King, B., McCracken, J., & State, M. (2014). Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(2), 237–257.

Wallis, K., & Pinto-Martin, J. (2008). The challenge of screening for autism spectrum disorder in a culturally diverse society. *Acta Paediatrica*, *97*(5), 539–540.

Wang, J., Hedley, D., Bury, S. M., & Barbaro, J. (2020). A systematic review of screening tools for the detection of autism spectrum disorder in mainland China and surrounding regions. *Autism*, *24*(2), 285–296.

Werner, E., Dawson, G., Munson, J., & Osterling, J. (2005). Variation in early developmental course in autism and its relation with behavioral outcome at 3–4 years of age. *Journal of Autism and Developmental Disorders*, *35*(3), 337–350.

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., ... Bossuyt, P. M. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, *155*(8), 529–536.

Woolfenden, S., Sarkozy, V., Ridley, G., & Williams, K. (2012). A systematic review of the diagnostic stability of Autism Spectrum Disorder. *Research in Autism Spectrum Disorders, 6*(1), 345–354.

*Wu, C. C., Chu, C. L., Stewart, L., Chiang, C. H., Hou, Y. M., & Liu, J. H. (2020). The utility of the screening tool for autism in 2-year-olds in detecting autism in Taiwanese toddlers who are less than 24 months of age: A longitudinal study. *Journal of Autism and Developmental Disorders*, *50*, 1172–1181.

Yuen, T., Penner, M., Carter, M. T., Szatmari, P., & Ungar, W. J. (2018). Assessing the accuracy of the modified checklist for autism in toddlers: A systematic review and meta-analysis. *Developmental Medicine & Child Neurology*, *60*(11), 1093–1100.

*Zahorodny, W., Shenouda, J., Mehta, U., Yee, E., Garcia, P., Rajan, M., & Goldfarb, M. (2018). Preliminary evaluation of a brief autism screener for young children. *Journal of Developmental and Behavioral Pediatrics*, *39*(3), 183–191.

*Zhou, H., Li, C., Luo, X., Wu, L., Huang, Y., Zhang, L., ... Wang, Y. (2018). Cross-cultural revision and psychometric properties of the Chinese version of the autism spectrum rating scale (2–5 years). *Frontiers in Neurology*, *9*, 460.

Zhou, W. Z., Ye, A. Y., Sun, Z. K., Tian, H. H., Pu, T. Z., Wu, Y. Y., ... Wei, L. (2014). Statistical analysis of twenty years (1993 to 2012) of data from mainland China's first

intervention center for children with autism spectrum disorder. *Molecular Autism, 5*(1), 1–14.

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience*, *23*(2-3), 143–152.

Zwaigenbaum, L., Bauman, M. L., Fein, D., Pierce, K., Buie, T., Davis, P. A., ... Wagner, S. (2015). Early screening of autism spectrum disorder: recommendations for practice and research. *Pediatrics, 136*(Suppl. 1), S41−S59.

Early screening tools for Autism Spectrum Disorder in infancy and toddlers

CHEN Guanghua[1], TAO Guanpeng[1], ZHAI Luyu[1], BAI Xuejun[2]

([1] *College of Preschool & Primary Education, Shenyang Normal University, Shenyang* 110034, *China*)

([2] *Academy of Psychology and Behaviour, Tianjin Normal University, Tianjin* 300074, *China*)

Abstract: The current study provides a systematic review of screening tools for the early detection of autism in infants and preschool children. A total of thirty-five empirical studies including infants (nine screening tools for a total of 159, 388 children in the sample), and toddlers (14 screening tools for 11, 712 children) met criteria for inclusion. The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) checklist was applied to provide a comprehensive understanding for the early identification of autism spectrum risk. The M-CHAT-R/F and PDQ-1 among all other screening tools were rated as good for infants; OERA and TIDOS were evaluated as excellent tools for young children. The classification accuracy of screening tools in early childhood was higher than that in infancy. The M-CHAT-R/F showed as one of the most promising measures. Lastly, we discussed the methodological limitations of QUADAS-2, and emphasized the importance of standardization of the evaluation of screening tools and the necessity of further validation studies for all the measures.

Key words: autism spectrum disorder, screening tools, the diagnostic sensitivity and specificity, QUADAS-2

附录

附录一：进行全文审查的质量评估标准及评分要求

评分的具体问题(注意事项)	圈出你的评估意见	
	Yes1	No0
1 是否明确描述(或引用)关于“筛查”为主题的测试?	1	0
2 是否以谱系患者为代表作为接受检测患者筛查的目的? (对于残疾儿童的研究，考虑是否可能代表较大的残疾人群，或者样本是否在某些方面存在显著差异)	1	0
3 是否明确描述了筛查被试的标准? 样本大小在大于 30 以上保留，如低于 30 以下可以备注删除	1	0
4 筛查的结果需要有明确的 ASD 诊断，诊断的参考标准是否符合正确分类目标条件? (诊断应包括以下内容之一: A. 基于 DSM-IV/DSM-5/ICD-10 的临床诊断+ADI-R+ADOS B. 基于 DSM-IV/DSM-5/ICD-10 的临床诊断+ADOS 和/或 ADI-R C. [基于 DSM-IV/DSM-5/ICD-10 的临床诊断+其他]或[ADOS/ADI-R+其他，如 SRS、CARS、SCQ、CAST、ASSQ 或 STAT、M-CHAT 等] D. 仅基于 DSM-IV/DSM-5/ICD-10 的临床诊断或仅 ADOS，或仅 ADI-R)	ABCD	
5 是否筛查的参考标准独立于测试? (筛查工具的名称，发明者与年代，工具的构成和评分标准，评分的要求和截止值，筛查程序，如质量标记、时间、成本等)	1	0
6 是否对执行筛查测试的过程描述得足够详细，以允许复制筛查测试? (明确地介绍了筛查的管理及过程性的内容，如：地点，筛查者，筛查时间，筛查人员的身份，专业背景、经过额外培训和学术水平等。)	1	0
7 当测试结果被解释为在实践中使用测试时，是否有相同的临床数据可用(筛查和诊断之间的时间间隔是否足够)?	1	0
8 是否解释了退出研究的原因(入组后)?	1	0
9 是否明确报告了筛查工具的心理学测量属性计算方法/数据? (如：评分者信度，内部一致性信度，分半信度，重测信度；效标效度，内容效度，结构效度等。)	1	0
10 是否明确报告了筛查工具的分类准确性计算方法/数据? (如：敏感性，特异性，阳性(PPV)，阴性(NPV)，至少要包括敏感性，特异性)	1	0

附录二：文中涉及的英文缩写之原文和中文翻译

ASD = 自闭症谱系障碍(Autism Spectrum Disorder)
GS = 普通样本(General Sample)
CS = 临床样本(Clinical Sample)
ATD =非典型性发展(atypically developing)
TD = 典型性发展(typically developing)
NCR=未提出任何问题(no concerns raised)
OPD = 其他精神障碍(Other Psychiatric Disorders)
LR = 低危 (low risk)
HR = 高危(high risk)
K = Kappa
A = 克伦巴赫 α 系数 Cronbach's α
ρ = Spearman 秩相关系数

NR = 未报告(Not Report)
ICD-9 = 《国际疾病分类》the International Classification of Disease-9 [世卫组织, 1992 年]
ICD-10 = 《国际疾病分类》the International Classification of Disease-10[世卫组织, 1992 年]
DSM-IV = 《精神障碍诊断和统计手册》第四版(the Diagnostic and Statistical Manual of Mental Disorders—fourth edition) [美国精神病学协会, 1994 年]
DSM-IV-TR = 《精神障碍诊断和统计手册》第四版修订版(the Diagnostic and Statistical Manual of Mental Disorders—fourth edition-text revision)[美国精神病学协会, 2000 年]
DSM-5 = 《精神障碍诊断和统计手册》第五版 (the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition)[美国精神病学协会, 2013 年]
ADI-R = 自闭症诊断访谈修订版(Autism Diagnostic Interview-Revised)
ADOS = 自闭症诊断观察表(Autism Diagnostic Observation Schedule)
DSM-PDP = 普遍性发展问题(Diagnostic and Statistical Manual of Mental Disorders-Pervasive Developmental Problems)
DISCO-11 = 社交和沟通障碍诊断访谈(the Diagnostic Interview for Social and Communication Disorders-11)

ABC = 自闭症行为检查表(Autism Behavior Checklist)
ASQ-3 = 年龄和阶段问卷第三版(Ages and Stages Questionnaires, Third Edition)
ASQ-SE = 年龄和阶段问卷：社会情绪(Ages and Stages Questionnaires: Social-Emotional)
ASRS = 自闭症谱系评定量表(the autism spectrum rating scale)
BASC-2 = 儿童行为评估系统-2(Behavioral Assessment System for Children-2)
Bayley III = BSITD-3 = 贝利婴幼儿发展量表第三版(Bayley Scales of Infant and Toddler Development, Third Edition)
BDI-2 = 巴特尔发展量表第 2 版(developmental profiles from the Battelle Developmental Inventory—2nd Edition)
CBCL/1½-5 = 1.5-5 岁儿童行为筛查表(Child behavior checklist)
CDI = 麦克阿瑟-贝茨交际发展量表(The MacArthur-Bates Communicative Development Inventory)
CARS-2 = 儿童孤独症评定量表第二版(Social Responsiveness Scale–Preschool)
CARS 2-ST = 儿童孤独症评定量表第二版，标准版(Social Responsiveness Scale–2 Standard)
CAT/CLAMS = 认知适应性测试/临床语言和听觉里程碑量表(the Cognitive Adaptive Test/Clinical Linguistic and Auditory Milestone Scale)
CPRS-R = 康纳斯父母评定量表修订版(the Conners' Parent Rating Scale-Revised)
Denver II = DDST-II = 丹佛发育筛查测试第二版 The Denver Developmental Screening Test II
DAS-II = 差异能力量表第二版(the Differential Ability Scales-Second Edition)
DCQ = 发展问题问卷(Developmental Concerns Questionnaire)
EADT = Enjoji 分析发展测验(Enjoji's Analytical Developmental Test)
GDS = 普遍性发育筛查 (Global Developmental Screening)
GMDS = 格里菲斯智力发展量表(the Griffiths Mental Development Scales)
MASC = 儿童多维焦虑量表(the Multidimensional Anxiety Scale for Children)
M-PRS = 美林-帕尔默修订的发展量表(Merrill-Palmer—Revised Scales of Development)
MSEL = 穆伦早期学习量表(Mullen Scales of Early Learning)
PEP-R = 心理教育概况修订版(Psychoeducational Profile-Revised)
SCAS = 斯彭斯儿童焦虑量表(the Spence Children's Anxiety Scale)

SON-R 2 ½-7 = SON-R 非语言智力测试(Snijders-Oomen Nonverbal Intelligence test)
SRS-P = 社会反应量表-学龄前儿童版本(the Social Responsiveness Scale–Preschool Version)
TASI = 幼儿自闭症症状访谈(Toddler Autism Symptom Interview)
TBIS = Tanaka 比奈智力量表(Tanaka Binet Intelligence Scale)
VABS = 文兰适应性行为量表(Vineland Adaptive Behavior Scale)
Vineland-II = Vineland 适应性行为量表第二版(the Vineland Adaptive Behaviour Scales—Second Edition)
WISC-III = 韦氏儿童智力量表第三版(Wechsler Intelligence Scale for Children, third edition)
WPPSI—III = 韦克斯勒学前和小学智力量表第三版(the Wechsler Preschool and Primary Scale of Intelligence—Third Edition)